```
OM of: US-09-303-518D-131 to: SwissProt_40:*   out_format : pfs

Date: Jun 30, 2002  8:35 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142714_4362/app_query.fasta_1.23501
 -DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09303518_@CGN1_1_440 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-131
Query length: 1344
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 217.960000


score_list:
Sequence                  Strd Orig     ZScore    EScore Len   ! Documentation  ..
SwissProt_40:NQRA_NEIMA + 2222.00 2930.16  3.6e-156  447 . ! Q9jvp8 neisseria meningitidi
SwissProt_40:NQRA_NEIMB + 2221.00 2928.83  4.3e-156  447 ! Q9k0m3 neisseria meningitidi
SwissProt_40:NQRA_PASMU + 1700.50 2240.55  9.4e-118  446  ! Q9clb1 pasteurella multocida
SwissProt_40:NQRA_HAEIN + 1646.00 2168.46  9.7e-114  447  ! P43955 haemophilus influenza
SwissProt_40:NQRA_VIBHA + 1491.50 1964.17  2.3e-102  446  ! Q9rfw1 vibrio harveyi. na(+)
SwissProt_40:NQRA_VIBAL + 1488.50 1960.20  3.9e-102  446  ! Q56586 v na(+)-translocating
SwissProt_40:NQRA_VIBCH + 1476.50 1944.33  3.0e-101  446  ! Q9kps1 vibrio cholerae. na(+
SwissProt_40:NQRA_PSEAE + 1358.00 1787.65  1.6e-92   445  ! Q9hzk6 pseudomonas aeruginos
SwissProt_40:NQRA_CHLMU +  453.00  590.48  7.3e-26   465  ! Q9plu3 chlamydia muridarum.
SwissProt_40:NQRA_CHLPN +  450.50  587.13  1.1e-25   467  ! Q9z7g2 chlamydia pneumoniae
SwissProt_40:NQRA_CHLTR +  432.00  562.71  2.6e-24   465  ! O84639 chlamydia trachomatis
SwissProt_40:RNFC_VIBCH +  172.00  214.17  4.0e-05   774  ! Q9kt88 vibrio cholerae. elec
SwissProt_40:YM96_YEAST -  143.00  172.23  0.0059   1140  ! Q04893 saccharomyces cerevi
SwissProt_40:RNFC_HAEIN +  135.50  165.38  0.0199    819  ! P71397 haemophilus influenza
SwissProt_40:RNFC_PSEAE +  132.00  161.27  0.0356    774  ! Q9hyb8 pseudomonas aeruginos
SwissProt_40:RNFC_BUCAI +  131.00  164.51  0.0385    473  ! P57215 buchnera aphidicola (
SwissProt_40:N121_RAT -  129.50  153.91  0.0590    1199 ! P52591 rattus norvegicus (rat
SwissProt_40:HKR1_YEAST -  127.00  146.83  0.0974   1802  ! P41809 saccharomyces cerevi
SwissProt_40:RNFC_RHOCA +  126.50  157.70  0.0839    519  ! Q52716 rhodobacter capsulatu
SwissProt_40:RNFC_ECO57 +  121.00  147.14  0.2280    740  ! P58324 escherichia coli o157
SwissProt_40:RNFC_ECOLI +  117.00  141.85  0.4494    740  ! P77611 escherichia coli. ele
SwissProt_40:ODP2_MYCCA +  116.50  146.05  0.4433    438  ! Q49110 mycoplasma capricolum
SwissProt_40:FAS1_SCHPO +  114.50  129.01  0.8330   2073  ! Q9uug0 s fatty acid synthas
SwissProt_40:PPSA_PYRAB +  113.00  135.62  0.9026    819  ! Q9v2h7 pyrococcus abyssi. pr
SwissProt_40:CHLN_SYNP7 +  112.50  140.18  0.8838    466  ! P54208 synechococcus sp. (st
SwissProt_40:PTN8_MOUSE -  112.00  134.49  1.07      802  ! P29352 mus musculus (mouse).
SwissProt_40:PPSA_PYRFU +  111.00  133.00  1.27      817  ! P42850 pyrococcus furiosus.
SwissProt_40:HFC1_MESAU -  109.50  122.32  1.95     2090  ! P51611 mesocricetus auratus
SwissProt_40:N121_HUMAN -  108.50  125.91  2.09     1229  ! Q9y2n3 homo sapiens (human)
SwissProt_40:MSB2_YEAST -  107.00  123.37  2.73     1306  ! P32334 saccharomyces cerevi
SwissProt_40:AMYH_YEAST -  107.00  122.94  2.75     1367  ! P08640 saccharomyces cerevi
SwissProt_40:MUC1_MOUSE -  106.50  129.46  2.59      630  ! Q02496 mus musculus (mouse).
SwissProt_40:ATPB_NICRU +  105.00  129.65  3.19      498  ! P26530 nicotiana rustica (az
SwissProt_40:ATPB_TOBAC +  105.00  129.65  3.19      498  ! P00826 nicotiana tabacum (cc
SwissProt_40:PPSA_PYRHO +  105.00  125.02  3.51      821  ! O57830 pyrococcus horikoshii
SwissProt_40:SMB2_MESAU +  105.00  123.30  3.63      989  ! Q60560 mesocricetus auratus
SwissProt_40:ILVE_BACSU +  104.50  131.94  3.27      362  ! P39576 bacillus subtilis. pu
SwissProt_40:MUC1_MESAU -  104.50  126.16  3.68      676  ! Q60528 mesocricetus auratus
SwissProt_40:MURF_HAEIN +  104.00  129.12  3.72      457  ! P45061 haemophilus influenza
SwissProt_40:SC18_SCHPO +  103.50  123.37  4.49      792  ! Q9p7q4 schizosaccharomyces p
SwissProt_40:VGP3_EBV -  103.50  122.12  4.61      907  ! P03200 epstein-barr virus (
SwissProt_40:TEGU_HCMVA +  103.50  113.74  5.46     2241  ! P16785 human cytomegalov
SwissProt_40:SGS3_DROER +  102.00  129.55  4.91      328  ! P13730 drosophila erecta
SwissProt_40:ATPB_NICPL +  102.00  125.68  5.31      498  ! P26529 nicotiana plumbagi
SwissProt_40:HUTH_BACSU +  101.50  124.84  5.80      508  ! P10944 bacillus subtilis.
SwissProt_40:IF2G_YEAST +  101.50  124.50  5.84      527  ! P32481 saccharomyces cere
SwissProt_40:APMU_PIG -  101.50  117.27  6.76     1150  ! P12021 sus scrofa (pig). a
SwissProt_40:RLA0_METJA +  101.00  127.95  5.85      338  ! P54049 methanococcus jann
SwissProt_40:YN96_YEAST -  101.00  116.88  7.32     1117  ! P53753 saccharomyces cer
SwissProt_40:DTXH_CORBE +  100.50  122.61  7.00      560  ! P00589 corynephage beta.
SwissProt_40:CHEB_PYRHO +  100.00  125.84  7.04      368  ! O58192 pyrococcus horikos
SwissProt_40:FIG2_YEAST -  100.00  112.18  9.29     1609  ! P25653 saccharomyces cer
SwissProt_40:ATPB_PYLLI +   99.50  122.70  8.06      481  ! P26532 pylaiella littoral
SwissProt_40:RNFC_PASMU +   99.50  117.59  8.94      835  ! Q9cnp2 pasteurella multoc
SwissProt_40:PPSA_AERPE +   99.50  117.48  8.96      845  ! Q9yec5 aeropyrum pernix.
SwissProt_40:RPOB_RICPR +   99.50  112.98  9.82     1374  ! O52271 rickettsia prowaz
SwissProt_40:DYHC_EMENI +   99.50  102.32  12.18    4344  ! P45444 emericella nidula
SwissProt_40:DTX_COROM +   98.50  119.97  9.83      560  ! P00587 corynephage omega.
SwissProt_40:DTX_CORBE +   98.50  119.85  9.85      567  ! P00588 corynephage beta. d
SwissProt_40:MKC7_YEAST +   98.50  119.39  9.95      596  ! P53379 saccharomyces cere
SwissProt_40:MLE_DROME -   98.00  111.56  12.52     1293  ! P24785 drosophila melanog
SwissProt_40:MURA_ECOLI +   97.50  121.33  11.03      419  ! P28909 escherichia coli.
SwissProt_40:VL2_HPV70 +   97.50  120.35  11.25      466  ! P50801 human papillomaviru
SwissProt_40:CHLN_SYNY3 +   97.50  120.29  11.27      469  ! P28372 synechocystis sp.
SwissProt_40:URE2_HELFE +   97.50  118.50  11.68      569  ! Q08716 helicobacter felis
SwissProt_40:YN26_MYCTU +   97.00  115.96  13.21      697  ! P71886 mycobacterium tube
SwissProt_40:DPO1_THEAQ +   97.00  114.32  13.65      832  ! P19821 thermus aquaticus.
SwissProt_40:LHS1_YEAST +   97.00  113.79  13.80      881  ! P36016 saccharomyces cere
SwissProt_40:FAS1_CANAL +   97.00  106.03  16.15     2037  ! P34731 c fatty acid synt
SwissProt_40:SRF2_BACSU +   96.50  100.13  19.55     3587  ! Q04747 bacillus subtilis
SwissProt_40:YG1F_YEAST -   96.00  116.81  14.98      551  ! P53214 saccharomyces cere
SwissProt_40:ANP_NOTCO -   96.00  113.48  16.02      790  ! P24856 notothenia coriicep
SwissProt_40:SMB2_MOUSE +   96.00  111.36  16.72      993  ! P40694 mus musculus (mous
SwissProt_40:SON_HUMAN +   96.00  103.09  19.78     2426  ! P18583 homo sapiens (huma
SwissProt_40:IMDH_PYRAB +   95.50  117.33  15.91      485  ! Q9uy49 pyrococcus abyssi.
SwissProt_40:NSF_CRIGR +   95.50  113.37  17.25      744  ! P18708 cricetulus griseus
SwissProt_40:YA54_METJA +   95.50  111.66  17.85      895  ! Q58454 methanococcus jann
SwissProt_40:BIOB_BACSH +   95.00  120.18  16.13      332  ! P19206 bacillus sphaericu
SwissProt_40:MS11_AGRRH -   95.00  117.79  16.94      430  ! P27874 agrobacterium rhiz
SwissProt_40:ATPB_DICDH +   95.00  116.75  17.30      481  ! P30158 dictyota dichotoma
SwissProt_40:ADEC_METTH +   95.00  115.71  17.66      538  ! O26952 methanobacterium t
SwissProt_40:AGA1_YEAST -   95.00  112.95  18.68      725  ! P32323 saccharomyces cere
SwissProt_40:HFC1_HUMAN -   95.00  103.39  22.67     2035  ! P51610 homo sapiens (hum
SwissProt_40:NSF_MOUSE +   94.50  112.05  20.43      744  ! P46460 mus musculus (mouse
SwissProt_40:GLYA_LACLA +   94.00  116.79  19.93      415  ! Q9chw7 lactococcus lactis
SwissProt_40:ATPB_ANASP +   94.00  115.41  20.50      482  ! P06540 anabaena sp. (stra
SwissProt_40:ATPB_NICSP +   94.00  115.10  20.63      498  ! P26531 nicotiana sp. (tob
SwissProt_40:PGMP_PEA +   94.00  112.99  21.53      626  ! Q9sm59 pisum sativum (garde
SwissProt_40:YNR6_YEAST -   94.00  112.84  21.60      636  ! P53882 saccharomyces cere
SwissProt_40:ICP0_HSVBJ +   94.00  112.27  21.84      676  ! P29128 bovine herpesvirus
SwissProt_40:MID2_YEAST -   93.50  117.04  21.30      376  ! P36027 saccharomyces cere
SwissProt_40:IF2G_SCHPO +   93.50  115.46  21.99      446  ! Q09130 schizosaccharomyce
SwissProt_40:AMPH_CHICK +   93.50  111.53  23.82      682  ! P50478 gallus gallus (chi
SwissProt_40:CTPV_MYCTU +   93.50  110.41  24.37      770  ! P77894 mycobacterium tube
SwissProt_40:EDD_ZYMMO +   93.00  112.32  25.17      583  ! P21909 zymomonas mobilis.
SwissProt_40:FSHR_MACFA +   93.00  110.70  26.02      695  ! P32212 macaca fasciculari
SwissProt_40:CARB_PSEAE +   93.00  106.68  28.22     1072  ! P38100 pseudomonas aerug
SwissProt_40:ALS1_CANAL +   93.00  105.19  29.09     1260  ! P46590 candida albicans
SwissProt_40:VGLX_HSVEB -   92.50  108.77  29.06      797  ! P28968 equine herpesvirus
SwissProt_40:VP6_BTV1S +   92.00  116.33  26.78      328  ! Q03328 bluetongue virus (s

seq_name: SwissProt_40:NQRA_NEIMA

seq_documentation_block:
ID   NQRA_NEIMA     STANDARD;      PRT;   447 AA.
AC   Q9JVP8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NQRA OR NMA0752.
```

```
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL162754; CAB84035.1; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW   Complete proteome.
SQ   SEQUENCE   447 AA;  48673 MW;  B808CAA82C901A1D CRC64;

alignment_scores:
            Quality:  2222.00          Length:    447
              Ratio:   4.982             Gaps:      0
  Percent Similarity:   99.776   Percent Identity: 96.197

alignment_block:
  US-09-303-518D-131 x NQRA_NEIMA

  Align seg 1/1  to: NQRA_NEIMA  from: 1  to: 447

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||||||||||||||||||||:::|||||||||||||||||||||||||
      17 uGlnValIleTyrAspGlyProValIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
         ||||||||:::||||||||||||||||||||||:::|||||||||:::|||
      34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
         ||||||||||||:::||||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProValSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

     301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
         ||||||||||||||:::||||||||||||||:::||||||:::|||::
     101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluGl 117

     351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
         :|||||||||||||||||||||||||||||||||||||||||||||||||
     117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
         ||||||||||||||||||||||||||||||||||||:::|||:::||
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProValValValIl 167

     501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
         ||||||||||||||||||||:::|||||||||||||||||||||||||
     167 eLysGluAlaAlaGluAspPheArgArgGlyLeuLeuValLeuSerArgL 184

     551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

     601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

     651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

     701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

     751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
         ||||||:::||||||||||||||||||||||||||||:::||||||||||
     251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyGl 267

     801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
         |  |||||||||||||||||||||||||||||||||||||||||||||||
     267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

     851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
         |||||||:::||||||||||||||||||||||||||||||||||||||||
     284 alSerGlnIleThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

     901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
         |||||||||||||||||||||||||:::||||||||||||||||||||
     301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317

     951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

    1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

    1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrThrAl 367

    1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

    1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400
```

```
1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
     |||||||||||||||||||||||||||||||||||||||||||||||||:
 417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434

1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
     |||||||||||||||||||||||||||||||||||||||||
 434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447
```

seq_name: SwissProt_40:NQRA_NEIMB

```
seq_documentation_block:
ID   NQRA_NEIMB     STANDARD;      PRT;   447 AA.
AC   Q9K0M3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NQRA OR NMB0569.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE002412; AAF40997.1; -.
DR   TIGR; NMB0569; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW   Complete proteome.
SQ   SEQUENCE   447 AA;  48635 MW;  BE174A78439BE477 CRC64;
```

```
alignment_scores:
           Quality: 2221.00           Length:    447
             Ratio:   4.980             Gaps:      0
Percent Similarity:  99.776  Percent Identity: 96.197
```

```
alignment_block:
 US-09-303-518D-131 x NQRA_NEIMB   ..

 Align seg 1/1  to: NQRA_NEIMB  from: 1  to: 447

   1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
   1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

  51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
     ||||:::::|||||||||||||||||||||||||||||||||||||||||
  17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

 101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
     |||||||||:::|||||||||||||||||||||:::||||||||:::|||
  34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50

 151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

 201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

 251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

 301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
     ||||||||||||||:::||||||||||||||||:::||||||:::|||::
 101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluGl 117

 351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
     :|||||||||||||||||||||||||||||||||||||||||||||||||
 117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

 401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

 451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

 501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

 551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

 601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

 651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

 701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
     |||||||||||||||||||||||||||||||||||||:::||||||||||
 234 snLysThrValTrpThrIleAsnTyrGlnAspValIleThrIleGlyArg 250

 751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
     ||||||:::|||||||||||||||||||||||||||:::|||||||||||
 251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyGl 267

 801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
     |   ||||||||||||||||||||||||||||||||||||||||||||||
 267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284
```

```
 851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
     ||||||||:::|||||||||||||||||||||:::|||||||||||||||
 284 alSerGlnIleThrAlaGlyGluLeuValAspThrAspAsnArgValIle 300

 901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
     |||||||||||||||||||||||||||:::||||||||||||||||||||
 301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317

 951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
     ||||||||||||||||||||||||||||||||||||||||||:::|||||
 351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367

1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434

1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
     |||||||||||||||||||||||||||||||||||||||||
 434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447
```

seq_name: SwissProt_40:NQRA_PASMU

```
seq_documentation_block:
ID   NQRA_PASMU     STANDARD;      PRT;   446 AA.
AC   Q9CLB1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NQRA OR PM1328.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE006171; AAK03412.1; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW   Complete proteome.
SQ   SEQUENCE   446 AA;  48168 MW;  5C5287CB77C4932D CRC64;
```

```
alignment_scores:
           Quality: 1700.50             Length:    447
             Ratio:  4.199                Gaps:      1
 Percent Similarity:  90.604   Percent Identity: 73.154

alignment_block:
 US-09-303-518D-131 x NQRA_PASMU   ..

 Align seg 1/1  to: NQRA_PASMU  from: 1  to: 446

   1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
     ||||||   |||||||||||||||||||||||||||:::|||   |||||
   1 MetIleThrIleLysLysGlyLeuAsnLeuProIleSerGlySerProGl 17

  51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
     ||||||||||   ||||||   ||||||||||||||||||||||||||||
  17 uGlnValIleArgAspGlyAsnAlaIleThrGluValAlaLeuLeuGlyG 34

 101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
     ||||||||||||||||||||||||||||||||::::::||||||::::::
  34 luGluTyrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50

 151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

 201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
     ||||||||||||||||||||||   ::::::|||||||||||||||   |
  67 lPheThrAlaProAlaSerGlyThrValThrAlaIleHisArgGlyAlaL 84

 251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
     ||||||||||||||||||||||||||   :::|||||||||::::::|||
  84 ysArgValLeuGlnSerValValIleLysIleGluGlyAsnGluGlnIle 100

 301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
        ||||||:::|||:::   |||   |||   :::|||:::||||||::
 101 ThrPheGluLysTyrThrThrGluGlnLeuAsnGlnLeuThrSerGluGl 117

 351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
     :||||||:::||||||      ||||||||||||||||||||||||||||
 117 nValArgGlnAsnLeuGlnAlaSerGlyLeuTrpThrAlaLeuArgThrA 134

 401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
     ||||||||||||||:::|||||||||||||||   |||   :::||||||
 134 rgProPheSerLysValProAlaValAspAlaThrProValSerIlePhe 150

 451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
     |||||||||||||||||||||||||||:::|||||||||   ||||||::
 151 ValAsnAlaMetAspThrAsnProLeuCysAlaAspProGlnValIleVa 167

 501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
     ::::::::::|||:::   |||:::   ||||||   |||||||||||||
 167 lGlnGlnSerAlaGlnAlaPheGluAlaGlyLeuThrValLeuSerArgL 184

 551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
     ||   |||   |||:::::::::||||||||||||:::|||::::::|||
 184 euHisGluGlyLysValTyrLeuCysLysAlaAlaAsnAlaSerIlePro 200
```

```
 601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
     |||   :::   |||||:::::::::   ||||||:::||||||||||||
 201 SerProSerIleAlaAsnLeuAspValLysGluPheAlaGlyProHisPr 217

 651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
     ||||||||||||||||||||||||||||||||||:::||||||:::|||:
 217 oAlaGlyLeuSerGlyThrHisIleHisPheIleAspProValSerAlaT 234

 701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
     ::|||    ||||||   ||||||||||||||||||||||||:::|||:::
 234 hrLysPheValTrpTyrIleAsnTyrGlnAspValIleAlaValGlyLys 250

 751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
     ||||||:::||||||:::|||:::::::::|||||||||:::|||:::||
 251 LeuPheThrThrGlyGluLeuAspValSerArgValValSerLeuAlaGl 267

 801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
     |   ||||||:::::|||||||||:::||||||||||||||||||||:::
 267 yProGlnValLysAsnProArgLeuValArgThrValLeuGlyAlaAsnL 284

 851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
     ::|||||||||||||||:::|||:::   |||::::::||||||||||||
 284 euSerGlnLeuThrAlaAsnGluValLysAspGlyGluAsnArgValIle 300

 901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
     ||||||||||||||:::||||||   |||   |||      ||||||||
 301 SerGlySerValLeuSerGlyAlaLysAlaAlaGlyProValAspTyrLe 317

 951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
     ||||||||||        |||:::||||||:::||||||||||||:::||||
 317 uGlyArgTyrAlaLeuGlnValSerValLeuGluGluGlyArgGluLysG 334

1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
     ||::::::||||||:::   |||      :::|||||||||::::::|||
 334 luPheLeuGlyTrpIleMetProGlyAlaAsnLysTyrSerLeuSerArg 350

1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
     |||:::||||||||||||   |||   ||||||||||:::||||||||||||
 351 ThrValLeuGlyHisPheSerLys...LysLeuPheAsnPheThrThrAl 366

1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
     |:::|||||||||:::||||||||||||||||||||||:::||||||||||
 366 aLeuAsnGlyGlyGluArgAlaMetValProIleGlyAlaTyrGluArgV 383

1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
     ||||||||||||||||:::||||||||||||||||||||||||    :::
 383 alMetProLeuAspIleIleProThrLeuLeuLeuArgAspLeuAlaAla 399

1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 400 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 416

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
     ||||||||||||||||:::||||||||||||||||||    ||||||||||
 416 uAspLeuAlaLeuCysThrPheValCysProGlyLysAsnGluTyrGlyP 433

1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
     ||||||||||::::::|||:::   |||||||||||||||
 433 roLeuLeuArgGlnAlaLeuAspLysIleGluLysGluGly 446
```

seq_name: SwissProt_40:NQRA_HAEIN

```
seq_documentation_block:
ID   NQRA_HAEIN     STANDARD;      PRT;   447 AA.
AC   P43955; P43956;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NQRA OR HI0164.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   REVISIONS.
RA   White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-75 FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RA   Barcak G.J., Heimer S.R.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   IDENTIFICATION AS NQR SYSTEM.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=96176316; PubMed=8601449;
RA   Hayashi M., Nakayama Y., Unemoto T.;
RT   "Existence of Na+-translocating NADH-quinone reductase in Haemophilus
RT   influenzae.";
RL   FEBS Lett. 381:174-176(1996).
RN   [5]
RP   IDENTIFICATION BY MASS SPECTROMETRY.
RX   MEDLINE=20137488; PubMed=10675023;
RA   Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA   Gray C., Fountoulakis M.;
RT   "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL   Electrophoresis 21:411-429(2000).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 63.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32702; AAC21836.1; -.
DR   EMBL; U20229; AAA62134.1; ALT_FRAME.
DR   TIGR; HI0164; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW   Complete proteome.
SQ   SEQUENCE   447 AA;  48884 MW;  4670C93FB5FF0912 CRC64;
```

```
alignment_scores:
           Quality: 1646.00           Length:    448
             Ratio:  4.146              Gaps:      2
 Percent Similarity:  88.616  Percent Identity: 70.089

alignment_block:
 US-09-303-518D-131 x NQRA_HAEIN   ..

 Align seg 1/1  to: NQRA_HAEIN  from: 1  to: 447

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         ||||||   ||||||||||||||:::|||||||||||||||:::|||
       1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyLysProAl 17

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
           |||||||::::::|||   |||::::::::|||||:::||||||
      17 aGlnValIleHisSerGlyAsnAlaValAsnGlnValAlaIleLeuGlyG 34

     101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
          |||||||||||||||||||||||||||||||::::::||||||::::::
      34 luGluTyrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||::
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValIl 67

     201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
         :||||||||||||||||||||     |||:::||||||:::|||||||||
      67 ePheThrAlaProAlaSerGlyThrIleThrAlaIleAsnArgGlyGluL 84

     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         |||||||||||||||||||||||||     ||||||||:::::::::|||
      84 ysArgValLeuGlnSerValValIleAsnValGluGlyAspGluLysIle 100

     301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
            |||   :::|||      |||   |||      ||||||||||||::
     101 ThrPheAlaLysTyrSerThrGluGlnLeuAsnThrLeuSerSerGluGl 117

     351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
         :|||::::::||||||||||:::||||||||||||||||||||||||||
     117 nValLysGlnAsnLeuIleGluSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||:::|||::::::::::::|||        :::||||||
     134 rgProPheSerLysValProSerIleGluSerGluAlaSerSerIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
         ||||||||||||||||||||||||||||||||||||||||||:::|||::::::
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProSerValValLe 167

     501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
         :||||||   ::::::||||||    :::|||||||    ||||||||||||
     167 uLysGluTyrSerGlnAspPheThrAsnGlyLeuThrValLeuSerArgL 184

     551 TGACCGAACGTAAA...ATCCATGTGTGTAAAGCAGCAGGCGCAGACGTG 597
          ||           |||    :::|||:::|||||||||:::    :::::::::
     184 euPheProSerLysProLeuHisLeuCysLysAlaGlyAspSerAsnIle 200

     598 CCGTCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCA 647
         |||:::    :::         |||::::::    |||:::|||    |||    ||
     201 ProThrAlaAspLeuGluAsnLeuGlnIleHisAspPheThrGlyValHi 217

     648 TCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCG 697
         ||||||||||||||     ||||||||||||||||||||||||:::|||||||||
     217 sProAlaGlyLeuValGlyThrHisIleHisPheIleAspProValGlyI 234

     698 CGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGA 747
            :::||||||||||||    |||||||||||||||||||||||||:::|||
     234 leGlnLysThrValTrpHisIleAsnTyrGlnAspValIleAlaValGly 250

     748 CGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGG 797
         :::||||||:::||||||:::|||    :::|||||||||::::::|||::
     251 LysLeuPheThrThrGlyGluLeuTyrSerGluArgValIleSerLeuAl 267

     798 CGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGA 847
         :|||   |||||::::::|||||||||:::|||||||:::::::||||||:
     267 aGlyProGlnValLysGluProArgLeuValArgThrThrIleGlyAlaA 284

     848 AGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTG 897
         :::::||||||||||||    :::||||||      :::    ||||||||
     284 snLeuSerGlnLeuThrGlnAsnGluLeuSerAlaGlyLysAsnArgVal 300

     898 ATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTA 947
         |||||||||||||||||    |||    ||||||:::    :::||||||||
     301 IleSerGlySerValLeuCysGlyGlnIleAlaLysAspSerHisAspTy 317

     948 TTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCA 997
         |||||||||||||          |||:::|||||||||    ||||||::::::|
     317 rLeuGlyArgTyrAlaLeuGlnValSerValIleAlaGluGlyAsnGluL 334

     998 AAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACG 1047
         ||||||:::||||||||||:::    ||||||    :::||||||||||:::|||
     334 ysGluPhePheGlyTrpIleMetProGlnAlaAsnLysTyrSerValThr 350

    1048 CGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGAC 1097
         ||||||:::|||||||||||||    |||    |||||||||:::||||||||
     351 ArgThrValLeuGlyHisPheSerLys...LysLeuPheAsnPheThrTh 366

    1098 AGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGC 1147
         |:::    |||||||||:::||||||||||||||||||||||||:::|||||||
     366 rSerGluAsnGlyGlyGluArgAlaMetValProIleGlySerTyrGluA 383

    1148 GCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATC 1197
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     383 rgValMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIle 399

    1198 GTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGA 1247
         ||||||||||||||||:::||||||||    ||||||||||||||||||||||||
     400 ValGlyAspThrAspGlyAlaGlnGluLeuGlyCysLeuGluLeuAspGl 416

    1248 AGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACG 1297
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     416 uGluAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrG 433

    1298 GCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
         ||    :::||||||:::|||||||:::    |||||||||||||||
     433 lySerIleLeuArgGlnValLeuAspLysIleGluLysGluGly 447

seq_name: SwissProt_40:NQRA_VIBHA

seq_documentation_block:
ID   NQRA_VIBHA     STANDARD;      PRT;   446 AA.
AC   Q9RFW1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NQRA.
OS   Vibrio harveyi.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=669;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BB120;
RX   MEDLINE=20056044; PubMed=10587447;
RA   Zhou W., Bertsova Y.V., Feng B., Tsatsos P., Verkhovskaya M.L.,
RA   Gennis R.B., Bogachev A.V., Barquera B.;
RT   "Sequencing and preliminary characterization of the Na+-translocating
RT   NADH:ubiquinone oxidoreductase from Vibrio harveyi.";
RL   Biochemistry 38:16246-16252(1999).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
```

```
CC      UBISEMIQUINONE TO UBIQUINOL.
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC      UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC      AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF165980; AAF15411.1; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport.
SQ   SEQUENCE   446 AA;  48365 MW;  9B05B38BB1D7A97C CRC64;


alignment_scores:
            Quality: 1491.50          Length:     447
              Ratio:  3.854             Gaps:       1
 Percent Similarity:  86.577   Percent Identity: 64.430

alignment_block:
 US-09-303-518D-131 x NQRA_VIBHA   ..

 Align seg 1/1  to: NQRA_VIBHA  from: 1  to: 446

      1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
        |||||    ||||||||||||||  ||||||||||||||    |||::
      1 MetIleThrIleLysLysGlyLeuAlaLeuProIleAlaGlyAlaProSe 17

     51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
        :||||||||   ||||||    :::|||    :::|||||||||||||||
     17 rGlnValIleAsnAspGlyLysThrIleLysLysValAlaLeuLeuGlyG 34

    101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
        |||||||||||||||||||||||:::|||    ::::::    |||:::
     34 luGluTyrValGlyMetArgProThrMetHisValArgValGlyAspGlu 50

    151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
        |||||||||:::||||||||||||||||||||||||||||||||||||||
     51 ValLysLysAlaGlnValLeuPheGluAspLysLysAsnProGlyValLy 67

    201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
         ||||||||||||||:::||||||:::          |||:::||||||    |
     67 sPheThrAlaProAlaAlaGlyLysValIleGluIleAsnArgGlyAlaL 84

    251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
        ||||||||||||||||||||||||    |||    |||::::::::::::
     84 ysArgValLeuGlnSerValValIleGluValAlaGlyGluGluGlnVal 100

    301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
             |||:::::::::              ||||||    |||:::    |||
    101 ThrPheAspLysPheGluAlaAlaGlnLeuAlaGlyLeuAspArgGluVa 117

    351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
        ::::::    :::||||:::::||||||||||||||||||||||||||||
    117 lIleLysThrGlnLeuValGluSerGlyLeuTrpThrAlaLeuArgThrA 134

    401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
        |||||||||||||:::||||||::::::::::            |||||||||
    134 rgProPheSerLysValProAlaIleGluSerAlaThrLysAlaIlePhe 150

    451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
        |||:::||||||||||||||||||||||||||||:::|||    :::||||||
    151 ValThrAlaMetAspThrAsnProLeuAlaAlaGlnProGluLeuIleIl 167

    501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
        |    |||        |||    |||        ||||||    :::||||||    |
    167 eThrGluGlnGlnGluAlaPheValAlaGlyLeuAspIleLeuSerAlaL 184

    551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
        |||||||    |||:::::||||||||     :::|||:::::::::|||
    184 euThrGluGlyLysValTyrValCysLys...SerGlyThrSerLeuPro 199

    601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
        |||::::::    :::|||:::|||    |||    |||    ||||||||||
    200 SerSerSerGlnSerAsnValGluGluHisValPheAspGlyProHisPr 216

    651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
        |||||||||:::|||||||||:::||||||:::    ||||||:::|||:
    216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyrProValAsnAlaG 233

    701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
        ::::::::::|||:::||||||||||||||||||||||||:::|||:::
    233 luAsnValAlaTrpSerIleAsnTyrGlnAspValIleAlaPheGlyGln 249

    751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
        ||||||:::||||||:::|||    |||:::|||||||||:::|||:::||
    250 LeuPheLeuThrGlyGluLeuTyrThrAspArgValValSerLeuAlaGl 266

    801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
        |        ||||||:::||||||||||||||||||||||:::||||||:::
    266 yProValValAsnAsnProArgLeuLeuArgThrValValGlyAlaSerL 283

    851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
        :::::::::||||||    :::|||||||:::    ::::::    |||||||||
    283 euGluAspLeuThrAspSerGluLeuMetProGlyGluValArgValIle 299

    901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
        |||||||||||||||:::|||:::    |||:::|||    |||    |||||
    300 SerGlySerValLeuSerGlyThrGlnAlaSerGlyProHisAlaTyrLe 316

    951 GGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
        ||||||||||||:::|||:::|||||:::::::|||||||||:::||||
    316 uGlyArgTyrHisGlnGlnValSerValLeuArgGluGlyArgAspLysG 333

   1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
        |||||||||||||||:::::|||        :::|||:::||||||||||:::
    333 luLeuPheGlyTrpAlaThrProGlyLysAsnLysPheSerIleThrLys 349

   1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
        :::    |||||||||::::::|||::::::|||||||::::::||||||::
    350 SerPheLeuGlyHisIlePheLysGlyGlnLeuPheAsnMetThrThrTh 366

   1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
        ::::||||||:::|||||||||||||||||||||||:::||||||||||
    366 rThrAsnGlySerAspArgAlaMetValProIleGlyAsnTyrGluArgV 383

   1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
        |||||||||||||||:::    ||||||||||||||||||||||    :::
    383 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuCysAla 399

   1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
        ||||||||||||||||||:::||||||:::||||||||||||||||||
    400 GlyAspThrAspSerAlaGlnThrLeuGlyAlaLeuGluLeuAspGluGl 416

   1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
        |||||||||||||||:::||||||||||||||||||||||||||||||
    416 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyrGluTyrGlyG 433

   1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
          |||||||||:::    |||:::||||||    |||||||||
    433 lnLeuLeuArgGluCysLeuAspThrIleValLysGluGly 446

seq_name: SwissProt_40:NQRA_VIBAL

seq_documentation_block:
ID   NQRA_VIBAL     STANDARD;      PRT;   446 AA.
AC   Q56586;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   (Na(+)-translocating NADH-quinone reductase subunit alpha) (Na(+)-
DE   translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NQRA OR NQR1.
OS   Vibrio alginolyticus.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=663;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=NCIMB 11038;
RX   MEDLINE=95104445; PubMed=7805867;
RA   Beattie P., Tan K., Bourne R.M., Leach D.R.F., Rich P.R., Ward F.B.;
RT   "Cloning and sequencing of four structural genes for the Na(+)-
RT   translocating NADH-ubiquinone oxidoreductase of Vibrio
RT   alginolyticus.";
RL   FEBS Lett. 356:333-338(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hayashi M., Unemoto T., Sugiyama A.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-10.
RX   MEDLINE=98149659; PubMed=9490015;
RA   Nakayama Y., Hayashi M., Unemoto T.;
RT   "Identification of six subunits constituting Na+-translocating NADH-
RT   quinone reductase from the marine Vibrio alginolyticus.";
RL   FEBS Lett. 422:240-242(1998).
RN   [4]
RP   SEQUENCE OF 1-9 AND 334-340.
RX   MEDLINE=95104444; PubMed=7805866;
RA   Hayashi M., Hirai K., Unemoto T.;
RT   "Cloning of the Na(+)-translocating NADH-quinone reductase gene from
RT   the marine bacterium Vibrio alginolyticus and the expression of the
RT   beta-subunit in Escherichia coli.";
RL   FEBS Lett. 356:330-332(1994).
RN   [5]
RP   INHIBITION OF ENZYMATIC ACTIVITY.
RX   MEDLINE=20016049; PubMed=10549856;
RA   Nakayama Y., Hayashi M., Yoshikawa K., Mochida K., Unemoto T.;
RT   "Inhibitor studies of a new antibiotic, korormicin, 2-n-heptyl-4-
RT   hydroxyquinoline N-oxide and Ag+ toward the Na+-translocating NADH-
RT   quinone reductase from the marine Vibrio alginolyticus.";
RL   Biol. Pharm. Bull. 22:1064-1067(1999).
RN   [6]
RP   REVIEW.
RX   MEDLINE=21145117; PubMed=11248187;
RA   Hayashi M., Nakayama Y., Unemoto T.;
RT   "Recent progress in the Na(+)-translocating NADH-quinone reductase
RT   from the marine Vibrio alginolyticus.";
RL   Biochim. Biophys. Acta 1505:37-44(2001).
RN   [7]
RP   REVIEW.
RX   MEDLINE=21145118; PubMed=11248188;
RA   Steuber J.;
RT   "Na(+) translocation by bacterial NADH:quinone oxidoreductases: an
RT   extension to the complex-I family of primary redox pumps.";
RL   Biochim. Biophys. Acta 1505:45-56(2001).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL.
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- ENZYME REGULATION: THIS REACTION IS TIGHTLY COUPLED TO THE NA(+)
CC       PUMPING ACTIVITY AND SPECIFICALLY REQUIRES NA(+) FOR ACTIVITY.
CC       INHIBITED BY KORORMICIN AND 2-N-HEPTYL-4-HYDROXYQUINOLINE N-OXIDE
CC       (HQNO).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF.
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z37111; CAA85476.1; -.
DR   EMBL; AB008030; BAA22910.1; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport.
FT   CONFLICT    337    337       W -> L (IN REF. 4).
SQ   SEQUENCE   446 AA;  48622 MW;  6D65ACAA53FE515C CRC64;


alignment_scores:
           Quality: 1488.50          Length:    447
             Ratio:  3.856             Gaps:      1
 Percent Similarity:  86.353   Percent Identity: 63.758

alignment_block:
 US-09-303-518D-131 x NQRA_VIBAL   ..

 Align seg 1/1  to: NQRA_VIBAL  from: 1  to: 446

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         ||||||   |||||||||||||||:::|||||||||||||||   |||::
       1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProSe 17

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         :|||||||||   ||||||   :::|||   :::||||||||||||||||
      17 rGlnValIleAsnAspGlyLysThrIleLysLysValAlaLeuLeuGlyG 34

     101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
         ||||||||||||||||||||||||:::|||    ::::::   |||:::
      34 luGluTyrValGlyMetArgProThrMetHisValArgValGlyAspGlu 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
         |||||||||:::||||||||||||||||||||||||||||||||||||
      51 ValLysLysAlaGlnValLeuPheGluAspLysLysAsnProGlyValLy 67

     201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
           |||||||||||||||:::|||||||:::       ::::::||||||   |
      67 sPheThrAlaProAlaAlaGlyLysValIleGluValAsnArgGlyAlaL 84

     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||   |||    |||::::::::::::
      84 ysArgValLeuGlnSerValValIleGluValAlaGlyGluGluGlnVal 100

     301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
            |||:::::::::            |||:::   |||:::   |||
     101 ThrPheAspLysPheGluAlaAlaGlnLeuSerGlyLeuAspArgGluVa 117

     351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
           ::::::   :::|||::::::||||||||||||||||||||||||||||
     117 lIleLysThrGlnLeuValAspSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||:::||||||:::::::::::::        |||||||||
     134 rgProPheSerLysValProAlaIleGluSerSerThrLysAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
         |||:::||||||||||||||||||||||||||||||||   |||   :::||||||
     151 ValThrAlaMetAspThrAsnProLeuAlaAlaLysProGluLeuIleIl 167

     501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
         |:::|||         |||    |||      ||||||   :::||||||    |
     167 eAsnGluGlnGlnGluAlaPheIleAlaGlyLeuAspIleLeuSerAlaL 184

     551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
         ||||||||    |||::::::|||||||||     :::|||::::::::::|||
     184 euThrGluGlyLysValTyrValCysLys...SerGlyThrSerLeuPro 199
```

```
 601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
       ::::::  :::|||:::|||   |||   |||   ||||||||||
 200 ArgSerSerGlnSerAsnValGluGluHisValPheAspGlyProHisPr 216

 651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
     |||||||||:::||||||||:::||||||:::    |||||:::|||:
 216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyrProValAsnAlaG 233

 701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
     :::::::::::|||:::||||||||||||||||||||||||:::|||:::
 233 luAsnValAlaTrpSerIleAsnTyrGlnAspValIleAlaPheGlyLys 249

 751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
     ||||||:::||||||:::|||    |||:::|||||||||:::|||:::||
 250 LeuPheLeuThrGlyGluLeuTyrThrAspArgValValSerLeuAlaGl 266

 801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
     |      ||||||:::||||||||||:::||||||||||:::||||||::::
 266 yProValValAsnAsnProArgLeuValArgThrValIleGlyAlaSerL 283

 851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
     ::::::::||||||   :::||||||:::    ::::::  |||||||||
 283 euAspAspLeuThrAspAsnGluLeuMetProGlyGluValArgValIle 299

 901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
     ||||||||||||||:::|||:::   |||   |||   |||   |||||
 300 SerGlySerValLeuThrGlyThrHisAlaThrGlyProHisAlaTyrLe 316

 951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
     ||||||||||||:::|||:::|||||||::::::||||||||||:::||||
 316 uGlyArgTyrHisGlnGlnValSerValLeuArgGluGlyArgGluLysG 333

1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
     |||||||||||||||:::    |||       :::|||:::|||:::||||||
 333 luLeuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 349

1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
     :::   |||||||||   :::|||::::::||||||::::::||||||:::
 350 SerPheLeuGlyHisValPheLysGlyGlnLeuPheAsnMetThrThrTh 366

1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
     ::::||||||:::||||||:::||||||||||||||||:::||||||||||
 366 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrGluArgV 383

1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
     |||||||||||||||:::   ||||||||||||||||||||||||   :::
 383 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuCysAla 399

1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||||||||||||||||||||||||||||:::||||||||||||||||
 400 GlyAspThrAspSerAlaGlnAlaLeuGlyAlaLeuGluLeuAspGluGl 416

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
     |||||||||||||||:::||||||||||||||||||||||||||||||||
 416 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyrGluTyrGlyT 433

1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
       ||||||||:::   |||:::||||||||||||||||||
 433 hrLeuLeuArgGluCysLeuAspThrIleGluLysGluGly 446
```

seq_name: SwissProt_40:NQRA_VIBCH

```
seq_documentation_block:
ID   NQRA_VIBCH     STANDARD;      PRT;   446 AA.
AC   Q9KPS1; Q9X4Q3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NQRA OR VC2295.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=99179036; PubMed=10077658;
RA   Haese C.C., Mekalanos J.J.;
RT   "Effects of changes in membrane sodium flux on virulence gene
RT   expression in Vibrio cholerae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:3183-3187(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL.
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF117331; AAD29962.1; -.
DR   EMBL; AE004300; AAF95439.1; ALT_INIT.
DR   TIGR; VC2295; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW   Complete proteome.
SQ   SEQUENCE   446 AA;  48624 MW;  428E8C397EBA163D CRC64;

alignment_scores:
            Quality: 1476.50          Length:    447
              Ratio:  3.845             Gaps:      1
 Percent Similarity:  85.906  Percent Identity: 62.416

alignment_block:
 US-09-303-518D-131 x NQRA_VIBCH   ..

 Align seg 1/1  to: NQRA_VIBCH  from: 1  to: 446

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         ||||||   |||||||||||||||:::|||||||||||||||   |||:::
       1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProSe 17

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         :|||||||||   ||||||   ||||||   :::||||||||||||||||
      17 rGlnValIleSerAspGlyLysAlaIleLysLysValAlaLeuLeuGlyG 34

     101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
         ||||||||||||||||||||||:::|||   ::::::   |||:::
      34 luGluTyrValGlyMetArgProThrMetHisValArgValGlyAspGlu 50
```

```
 151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
     ||||||||:::|||:::||||||||||||||||||||||||||||
  51 ValLysLysAlaGlnIleLeuPheGluAspLysLysAsnProGlyValLy 67

 201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
      |||||:::|||:::||||||||:::::: |||:::||||||   |
  67 sPheThrSerProValSerGlyLysValValGluIleAsnArgGlyAlaL 84

 251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
     |||||||||||||||||||||||||   |||   |||:::|||::::::
  84 ysArgValLeuGlnSerValValIleGluValAlaGlyAspAspGlnVal 100

 301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
        |||:::::::::      :::  ||||||:::|||:::   :::
 101 ThrPheAspLysPheGluAlaAsnGlnLeuAlaSerLeuAsnArgAspAl 117

 351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
      ::::::   :::|||::::::||||||||||||||||||:::|||||||
 117 aIleLysThrGlnLeuValGluSerGlyLeuTrpThrAlaPheArgThrA 134

 401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
     ||||||||||||||:::||||||:::|||:::        |||||||||
 134 rgProPheSerLysValProAlaIleAspSerThrSerGluAlaIlePhe 150

 451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
     |||:::||||||||||||||||||||||||||||:::||||||||||:::||
 151 ValThrAlaMetAspThrAsnProLeuAlaAlaGluProThrValValIl 167

 501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
     |:::|||   :::|||  |||     ||||||   |||||||||   |
 167 eAsnGluGlnSerGluAlaPheValAlaGlyLeuAspValLeuSerAlaL 184

 551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
     |||||       |||::::::|||||||||       |||:::::::::|||
 184 euThrThrGlyLysValTyrValCysLys...LysGlyThrSerLeuPro 199

 601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
       ::::::       |||:::|||   |||   |||   |||||||||||
 200 ArgSerGlnGlnProAsnValGluGluHisValPheAspGlyProHisPr 216

 651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
     ||||||||||:::|||||||||:::|||||||:::   ||||||:::|||:
 216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyrProValSerAlaA 233

 701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
     ::    ::::::|||:::|||||||||||||||||||||||:::|||:::
 233 spHisValAlaTrpSerIleAsnTyrGlnAspValIleAlaValGlyGln 249

 751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
     ||||||:::||||||:::|||   |||:::|||||||||:::|||:::||
 250 LeuPheLeuThrGlyGluLeuTyrThrGlnArgValValSerLeuAlaGl 266

 801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
     |       ||||||||||||||||||:::||||||||||:::||||||:::
 266 yProValValAsnLysProArgLeuValArgThrValMetGlyAlaSerL 283

 851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
     :::::||||||:::    :::|||::::::   ::::::   |||:::|||
 283 euGluGlnLeuValAspSerGluIleMetProGlyGluValArgIleIle 299

 901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
     |||||||||||||||:::|||:::   |||   |||   |||   |||||
 300 SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaTyrLe 316

 951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
     ||||||||||||   |||:::||||||::::::||||||||||:::||||
 316 uGlyArgTyrHisLeuGlnValSerValLeuArgGluGlyArgAspLysG 333

1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
     |||||||||||||||:::   |||       :::|||:::|||:::||||||
 333 luLeuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 349

1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
     :::   |||||||||::::::|||:::::::::::::::::::||||||:::
 350 SerPheLeuGlyHisLeuPheLysGlyGlnValTyrAsnMetThrThrTh 366

1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
     ::::||||||:::||||||:::|||||||||||||||:::||||||:::|
 366 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrGluLysV 383

1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
     ||||||||||||||:::   |||||||||||||||||||||   :::
 383 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuCysAla 399

1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||:::|||||||||       ||||||:::|||||||||||||||
 400 GlyAspSerAspSerAlaValArgLeuGlyAlaLeuGluLeuAspGluGl 416

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
     ||||||||||||||||:::|||||||||||||||||||||||||||||
 416 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyrGluTyrGlyG 433

1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
        ||||||||:::   |||:::   |||||||||||||||
 433 lnLeuLeuArgGluCysLeuAspLysIleGluLysGluGly 446
```

```
seq_name: SwissProt_40:NQRA_PSEAE

seq_documentation_block:
ID   NQRA_PSEAE     STANDARD;      PRT;   445 AA.
AC   Q9HZK6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NQRA OR PA2999.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE004724; AAG06387.1; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
```

```
KW   Complete proteome.
SQ   SEQUENCE   445 AA;  48081 MW;  ED0468D06559B701 CRC64;

alignment_scores:
            Quality: 1358.00          Length:    447
              Ratio:  3.641             Gaps:      1
 Percent Similarity:  83.445   Percent Identity: 59.060

alignment_block:
 US-09-303-518D-131 x NQRA_PSEAE   ..

 Align seg 1/1  to: NQRA_PSEAE  from: 1  to: 445

      1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
        ||||||||||||||:::|||||:::||||||||||:::|||    |||||
      1 MetIleLysIleLysArgGlyLeuAspLeuProIleSerGlyAlaProGl 17

     51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
        ||||  |||        :::      :::   :::||||||||||::::|||
     17 uGlnArgIleGluAlaAlaArgProValArgSerValAlaLeuIleGlyP 34

    101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
          :::|||   ||||||:::|||:::|||   ::::::   |||:::
     34 heAspTyrHisGlyMetLysProThrMetAlaValGlnValGlyAspArg 50

    151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
        ||||||   ||||||||||||||   ||||||||||||||||:::|||
     51 ValLysLeuGlyGlnValLeuPheThrAspLysLysAsnProSerValSe 67

    201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
         :::||||||||::::::|||   ::::::||||||||||||||||||||
     67 rTyrThrAlaProGlyAlaGlyValValSerAlaIleHisArgGlyGluL 84

    251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
        ||||||||||||||||||||||||   ::::::|||::::::::::::::
     84 ysArgValLeuGlnSerValValIleAspLeuAspGlyAspGluGlnLeu 100

    301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
        ||||||  ||||||       :::   ||||||   |||||||:::|||::
    101 GluPheAlaArgTyrProAlaAspLysLeuAlaThrLeuSerAlaGluGl 117

    351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
        :||||||  ||||||:::||||||||||||||||||||||||||||||||
    117 nValArgAspAsnLeuLeuGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

    401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
        |||||||||||||||:::|||      ::::::::::|||   :::||||||
    134 rgProPheSerLysValProAspProGluSerSerProSerSerIlePhe 150

    451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
        |||:::|||:::||||||:::|||||||||||||||||   |||:::||
    151 ValThrAlaIleAspThrGlnProLeuAlaAlaAspProGlnValValIl 167

    501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
        |  |||   :::|||   |||:::   ||||||   ||||||:::||||
    167 eAlaGluGlnGlyGluAlaPheGlnAlaGlyLeuThrValLeuGlyArgL 184

    551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
        |||:::        ::::::   :::|||||||||   |||:::::::::|||
    184 euAla......ArgValPheLeuCysLysAlaGluGlyValSerLeuPro 198

    601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
        :::|||     :::::::::   ::::::    |||:::|||||||||||||
    199 GlyGluAlaLeuSerGlyValThrAlaGlnAlaPheSerGlyProHisPr 215

    651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
        ||||||||||   ||||||||||||||||||||::::::||||||||||||||:
    215 oAlaGlyLeuProGlyThrHisIleHisPheLeuAspProValGlyAlaG 232

    701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
        ::|||:::||||||:::::|||||||||||||||||||||||||||:::
    232 lyLysSerValTrpAsnLeuAsnTyrGlnAspValIleAlaIleGlyLys 248

    751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
        ||||||:::||||||:::|||   |||||||||||:::||||||:::||
    249 LeuPheThrThrGlyGlnLeuTrpThrGluArgValIleAlaLeuAlaGl 265

    801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
        |       |||:::||||||||||||:::||||||   ||||||||||:::
    265 yProValValGluLysProArgLeuValArgThrArgLeuGlyAlaAsnL 282

    851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
        ::::::::|||:::||||||:::|||      ::::::||||||:::|||
    282 euAspGluLeuAlaAlaGlyGlnLeuGlnProGlyAsnAsnArgLeuIle 298

    901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
        |||||||||||||||:::|||      |||:::||||||:::     |||||
    299 SerGlySerValLeuGlyGlyArgThrAlaHisGlyAlaTyrAlaTyrLe 315

    951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
        ||||||||||||   |||:::|||    ::::::||||||    ::::::|
    315 uGlyArgTyrHisLeuGlnLeuSerCysLeuLysGluGlyAspGlnArgG 332

   1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
        ||::::::   ::::::             :::|||:::|||:::    :::
    332 luPheLeuHisTyrLeuArgAlaGlyValGluLysHisSerLeuLeuAsn 348

   1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
        :::    ::::::::::::|||    :::|||    |||    |||||||||::
    349 ValPheValSerArgLeuLeuGlyGlyLysArgPheAlaPheThrThrSe 365

   1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
        ::::||||||:::    |||||||||||||||:::|||:::||||||    |
    365 rThrAsnGlySerProArgAlaMetValProValGlyAsnTyrGluAlaV 382

   1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
        |||||||||||||||||||||||||   |||||||||   |||||||||
    382 alMetProLeuAspIleLeuProThrGlnLeuLeuArgTyrLeuIleVal 398

   1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
        |||||||||:::   ||||||   ||||||:::|||||||||||||||||
    399 GlyAspThrGluMetAlaGlnLysLeuGlyAlaLeuGluLeuAspGluGl 415

   1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
        ||||||||||||||||||:::||||||   ||||||||||||||||||
    415 uAspLeuAlaLeuCysSerTyrValCysAlaGlyLysTyrGluTyrGlyP 432

   1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
        ||:::||||||         |||      ||||||:::||||||
    432 roIleLeuArgAspAsnLeuAlaArgIleGluGlnGluGly 445

seq_name: SwissProt_40:NQRA_CHLMU

seq_documentation_block:
ID   NQRA_CHLMU     STANDARD;      PRT;   465 AA.
AC   Q9PLU3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable Na(+)-translocating NADH-quinone reductase subunit A
DE   (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)
DE   (NQR complex subunit A) (NQR-1 subunit A).
GN   NQRA OR TC0002.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MOPN / NIGG;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
```

```
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE002269; AAF38896.1; -.
DR   TIGR; TC0002; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW   Complete proteome.
SQ   SEQUENCE   465 AA;  51588 MW;  12A11ABE6FA64F2C CRC64;

alignment_scores:
           Quality:  453.00            Length:    464
             Ratio:  1.520               Gaps:     14
 Percent Similarity:  64.224   Percent Identity: 28.448

alignment_block:
 US-09-303-518D-131 x NQRA_CHLMU   ..

 Align seg 1/1  to: NQRA_CHLMU  from: 1  to: 465

        4 ATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCA 53
          |||   :::::::::||||||:::|||    :::    |||    |||:::::
        3 IleIleValSerArgGlyLeuAspLeuSerLeuLysGlyAlaProLysGl 19

       54 AGTCATTTATGACGGC.........CCGGCCATTACCGAAGTCGCGTTGC 94
          :       :::   |||          ||||||:::::::::|||   |||
       19 uSerGlyPheCysGlyLysValAspProAlaPheValSerValAspLeu. 35

       95 TTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGT 144
                :::   ::::::    :::        ::::::|||::::::    |||
       36 .....ArgProPheAlaProLeuProLeuGlyValLysValSerProGly 50

      145 GAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGG 194
          :::   :::      |||:::   |||   |||   |||:::   |||||
       51 AspGlnIleThrAlaGlySerProLeuAlaGluTyrLysSerPheProGl 67

      195 CGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTG 244
          ||||    :::|||:::    ::::::|||::::::        |||:::||||
       67 yValPheIleThrSerSerValAspGlyGluValIleGluIleArgArgG 84

      245 GCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAA...GGCAAC 291
          ||:::|||||||:::|||    ::::::||||||        :::    |||
       84 lySerLysArgAlaLeuLeuAspIleValIleLysLysLysProGlyVal 100

      292 GACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAG 341
          ::::::   :::|||:::                          |||      |||||
      101 SerGlnThrLysPheSer............TyrAspLeuHisAlaLeuSe 113

      342 CAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGC 391
          |::::::::::::    :::    :::    ::::::|||||||:::|||       :
      113 rGlnLysGluLeuLeuGluValPheLysLysGluGlyLeuPheThrLeuP 130

      392 TTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAG...CCG 438
          :::::    ||||||||||:::    |||||||||:::    ::::::    |||

      130 heLysGlnArgProPheAsn...IleProAlaLeuProThrGlnSerPro 145

      439 TTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTG......GCTGC 482
              :::|||:::|||        |||::::::|||:::          :::::
      146 ArgAspValPheIleAsnLeuAlaAspAsnArgProPheThrProSerVa 162

      483 CGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGAC...........T 520
          ::::        :::    :::    ::::::    ||||||                |
      162 lGluLysHisLeuSerLeuPheSerSerLysGluAspGlyTyrTyrIleP 179

      521 TCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCAT 570
          ||       |||:::    ::::::::::::|||           |||     |||
      179 heValValGlyValGlnAlaIleAlaLysLeuPheGlyLeuLysProHis 195

      571 GTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATAT 620
          :::    ::::::                   ::::::|||::::::::::          :::::
      196 IleValSerThrAspArgLeuSerLeuProThrGlnAspLeuIleSerVa 212

      621 C...GAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCA 667
          :    :::    |||    ::::::|||||||:::|||:::|||          :::|
      212 lAlaHisLeuHisThrIleAlaGlyProTyrProSerGlySerProSerT 229

      668 CGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAA...ACCGTGTGG 714
          ||||||||||    |||         :::         ::::::        |||:::
      229 hrHisIleHisHisIleAlaArgIleArgAsnAspArgAspIleValPhe 245

      715 ACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGG 764
          ||||||:::::|||:::|||::::::||||||:::||||||:::       ||
      246 ThrIleSerPheGlnGluValLeuSerIleGlyHisLeuPheLeuLysGl 262

      765 CCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGC......CTGCAAG 808
          |    :::          ::::::||||||||||||:::|||          |||
      262 yPhePheLeuGlyGlnGlnValValAlaLeuAlaGlySerAlaLeuProP 279

      809 TCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAA 858
           ::::::    :::    |||    |||:::    ||||||:::    ::::::
      279 roSerGlnArgLysTyrLeuIleThrAlaLysGlyAlaSerPheLysAsp 295

      859 CTTACCGCCGGCGAATTGGTTGACGCG...GACAACCGCGTGATTTCCGG 905
          |||          |||:::    ::::::       |||          :::||||||||
      296 LeuLeuProGlnGluIlePheSerSerAsnAspValSerLeuIleSerGl 312

      906 TTCGGTATTGAACGGTGCGATTGCA...CAAGGCGCGCATGATTATTTGG 952
          |:::    |||:::|||    ::::::    :::          :::       ||||
      312 yAspProLeuThrGlyArgLeuCysAsnLysGluGluAsnProCysLeuG 329

      953 GACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAG 1002
          ||          :::    |||:::::::::    :::    :::::::::|||
      329 lyMetArgAspHisThrIleThrIleLeuProAsnProLysThrArgGlu 345

     1003 CTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCAC 1052
          :::|||:::::::::                    :::|||    ::::::|||||||
      346 MetPheSerPheLeuArgLeuGlyTrpAsnLysLeuThrValThrArgTh 362

     1053 CACTCTCGGCCATTTCCTAAAAAACAAA......CTCTTCAAGTTCACGA 1096
          |    |||:::    |||:::|||:::|||           ::::::    ::::::|
      362 rTyrLeuSerGlyPhePheLysArgLysArgValPheMetAspMetAsnT 379

     1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAG 1146
          ||    ::::::|||          |||    ::::::                 ||||||
      379 hrAsnLeuHisGlyGluLysArgProIleIleAspSerGluIleTyrGlu 395

     1147 CGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAAT 1196
          :::|||          :::    :::          :::    |||::::::    |||
      396 LysValSerAlaIleAlaValProValAlaProLeuIleLysAlaLeuGl 412

     1197 CGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACG 1246
           :::    :::    ::::::|||          ||||||    ||||||::::::
      412 uThrGlnAsnPheGluGluAlaCysArgLeuGlyLeuLeuGluValSerP 429

     1247 AAGAAGACCTCGCTTTGTGCAGCTTCGTCTCCCCGGGCAAATACGAATAC 1296
           ||||||:::||||||    :::|||:::    |||:::|||    |||
      429 roGluAspPheAlaLeuProThrPheIleAspProSerLysThrGluMet 445
```

```
    1297 GGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAA 1338
                ::::::::::::::|||           ||||||
     446 PheAlaIleValLysGluAlaLeuIleArgTyrAlaLysGlu 459

seq_name: SwissProt_40:NQRA_CHLPN

seq_documentation_block:
ID   NQRA_CHLPN     STANDARD;      PRT;   467 AA.
AC   Q9Z7G2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable Na(+)-translocating NADH-quinone reductase subunit A
DE   (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)
DE   (NQR complex subunit A) (NQR-1 subunit A).
GN   NQRA OR NQR1 OR CPN0743 OR CP0002.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE001656; AAD18882.1; -.
DR   EMBL; AE002164; AAF37899.1; -.
DR   EMBL; AP002547; BAA98950.1; -.
DR   TIGR; CP0002; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW   Complete proteome.
SQ   SEQUENCE   467 AA;  52266 MW;  84CE73278BA0BFD6 CRC64;

alignment_scores:
          Quality:  450.50          Length:    470
            Ratio:  1.532             Gaps:     16
 Percent Similarity:  62.553  Percent Identity: 29.149

alignment_block:
 US-09-303-518D-131 x NQRA_CHLPN   ..

 Align seg 1/1  to: NQRA_CHLPN  from: 1  to: 467

       4 ATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCG...GA 50
         |||   ::::::::||||||:::|||   :::   |||   |||   ||
       3 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysGl 19

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         |:::   :::|||:::                         :::
      19 uSerGlyPheTyrAsn.......................LysIleAspP 28

     101 AAGAATATGTCGGC......ATGCGCCCC..................TCG 126
           |||:::|||:::      :::||||||                 :::
      28 roGluPheValSerIleAspLeuArgProPheGlnProLeuSerLeuLys 44

     127 ATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGA 176
         :::|||:::::::::|||:::||||||   :::|||      :::   ||
      45 LeuLysValGluGlnGlyAspAlaValCysSerGlyAlaProIleAlaGl 61

     177 AGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAA 226
         |   |||         |||::::::   :::|||:::   :::||||||   :
      61 uTyrLysHisPheProAsnThrTyrIleThrSerHisValSerGlyValV 78

     227 TCGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATT 276
         :::::||||||:::||||||:::||||||   |||   :::|||:::|||
      78 alThrAlaIleArgArgGlyAsnLysArgSerLeuLeuAspValIleIle 94

     277 ...GCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGA 323
               :::   |||      :::   |||:::      |||      ::
      95 LysLysThrProGlyProThrSerThrGluTyr...ThrTyrAspLeuGl 110

     324 AGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAAT 373
         ::::|||::::::   :::   :::::::::   :::         ::::
     110 nThrLeuSerArgSerAspLeuSerGluIlePheLys.........GluA 124

     374 CAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCC 423
         ::|||||||:::::   ::::::   ||||||||||:::   |||||||||
     124 snGlyLeuPheAlaLeuIleLysGlnArgProPheAsp...IleProAla 139

     424 GTAGATGCCGAG...CCGTTCGCCATCTTCGTCAATGCGATGGACACCAA 470
         :::   ::::::   |||      :::|||:::|||      |||:::::
     140 IleProThrGlnThrProArgAspValPheIleAsnLeuAlaAspAsnAr 156

     471 TCCGCTGGCTGCCGACCCTACGGTCATCATC..................A 502
         :|||::::::   :::|||         :::                  :
     156 gProPheThrProSerProGluLysHisLeuAlaLeuPheSerSerArgG 173

     503 AAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTG 552
         ::|||:::            |||      |||:::   ::::::::::|||
     173 luGluGlyPheTyrValPheValValGlyValArgAlaIleAlaLysLeu 189

     553 ACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTC 602
               :::   |||:::                     :::|||::
     190 PheGlyLeuArgProHisIleValPheArgAspArgLeuThrLeuProTh 206

     603 TGAAAATGCTGCCAATATC...GAAACACATGAATTTGGCGGCCCGCATC 649
         ::::::      :::|||   :::   |||      :::||||||   |
     206 rGlnGluLeuLysThrIleAlaHisLeuHisThrValSerGlyProPheP 223

     650 CTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCG 699
         ||:::|||      :::   |||||||||   :::   |||:::
     223 roSerGlySerProSerIleHisIleHisSerValAlaProIleThrAsn 239
```

```
 700 AATAAA...ACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGG 746
     :::|||   :::|||:::|||::::::::||||||||||::::::|||||
 240 GluLysGluValValPheThrLeuSerPheGlnAspValLeuThrIleGl 256

 747 ACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGG 796
     |:::|||||||:::   ||||||:::       |||:::|||:::||||||:
 256 yHisLeuPheLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuA 273

 797 GCGGC......CTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840
     ::|||       |||:::   :::       |||   :::   |||:::
 273 laGlyThrAlaLeuLysSerSerLeuArgArgTyrValIleThrThrLys 289

 841 GGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAA 890
     ||||||:::   |||:::|||             :::   ::::::   :::::
 290 GlyAlaSerPheSerSerLeuIleAsnLeuAsnAspIleSerAspAsnAs 306

 891 CCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGC 940
     :   :::|||||||||:::   |||:::|||   ::::::::         :
 306 pThrLeuIleSerGlyAspProLeuThrGlyArgLeuCysLysLysGluG 323

 941 ATGAT...TATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAA 987
     :::::   :::||||||          ::::::||||||||||:::::::::
 323 luGluProPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsn 339

 988 GGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATA 1037
        ::::::|||||||||:::::::::                 :::|||
 340 ProThrLysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysPr 356

1038 CTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCA 1087
     ::::::|||:::|||   |||:::   |||:::|||:::|||
 356 oThrPheThrLysThrTyrLeuSerGlyPhePheLysLysLysArg...T 372

1088 AGTTCACG.........ACAGCCGTCAACGGCGGCGACCGCGCCATGGTA 1128
     :::|||          |||   :::::::|||       |||   ::::::
 372 hrTyrThrAsnProAspThrAsnLeuHisGlyGluThrArgProIleIle 388

1129 CCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTT 1178
                |||:::::::||||||||||:::   |||       :::
 389 AspThrAspIleTyrAspLysValMetProMetArgIleProValValPr 405

1179 GCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGG 1228
     |||::::::   :::|||:::   :::   |||   |||:::   ||||
 405 oLeuIleLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuG 422

1229 GTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGC 1278
     ||   |||||||:::      |||||:::||||||   :::::::::
 422 lyPheLeuGluValCysGlyGluAspPheAlaLeuProThrLeuIleAsp 438

1279 CCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCAT 1328
     |||:::|||   |||        :::::::::::   |||
 439 ProSerLysThrGluMetLeuThrIleValLysGluSerLeuIleGluTy 455

1329 TGAGAAGGAA 1338
        ||||||
 455 rAlaLysGlu 458
```

seq_name: SwissProt_40:NQRA_CHLTR

seq_documentation_block:

```
ID   NQRA_CHLTR     STANDARD;      PRT;   465 AA.
AC   O84639;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable Na(+)-translocating NADH-quinone reductase subunit A
DE   (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)
DE   (NQR complex subunit A) (NQR-1 subunit A).
GN   NQRA OR NQR1 OR CT634.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC       UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC       OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC       ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC       UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC       UBIQUINOL + NA(+)(OUT).
CC   -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC       AND NQRF (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE001334; AAC68238.1; -.
KW   Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW   Complete proteome.
SQ   SEQUENCE   465 AA;  51757 MW;  303F88DC133AB998 CRC64;
```

alignment_scores:

```
          Quality:  432.00          Length:    464
            Ratio:  1.474             Gaps:     14
 Percent Similarity:  63.147   Percent Identity: 28.017
```

alignment_block:

```
 US-09-303-518D-131 x NQRA_CHLTR   ..

 Align seg 1/1  to: NQRA_CHLTR  from: 1  to: 465

   4 ATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCA 53
     |||   :::::::::||||||:::|||      :::   |||   |||::::::
   3 IleValValSerArgGlyLeuAspLeuSerLeuLysGlyAlaProLysGl 19

  54 AGTCATTTATGACGGC.........CCGGCCATTACCGAAGTCGCGTTGC 94
     :      :::   |||          |||:::   ::::::|||   |||
  19 uSerGlyPheCysGlyLysValAspProThrTyrValSerValAspLeu. 35

  95 TTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGT 144
             :::   ::::::   :::      ::::::|||:::
  36 .....ArgProPheAlaProLeuProLeuGlyValLysValThrProGlu 50

 145 GAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGG 194
     :::   |||      |||:::   |||   |||   |||          ||
  51 AspGlnValThrAlaGlySerProLeuAlaGluTyrLysLeuPheSerGl 67

 195 CGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTG 244
     ||||   :::|||:::|||::::::|||:::::::::   |||:::||||
  67 yValPheIleThrSerProValAspGlyGluValValGluIleArgArgG 84

 245 GCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAA...GGCAAC 291
     ||:::||||||:::|||   ::::::||||||       :::   |||
  84 lyAsnLysArgAlaLeuLeuGluIleValIleLysLysLysProGlyIle 100

 292 GACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAG 341
     ::::::   :::|||:::                  |||   :::|||::
 101 SerGlnThrLysPheSer............TyrAspLeuGlnSerLeuTh 113

 342 CAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGC 391
```

```
        :::::::   :::   :::   :::   ::::::|||||||::::::   :
 113 rGlnLysAspLeuLeuGluValPheLysLysGluGlyLeuPheAlaLeuP 130

 392 TTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAG...CCG 438
     :::::   |||||||||:::   |||||||||:::   ::::::   |||
 130 heLysGlnArgProPheAsp...IleProAlaLeuProThrGlnSerPro 145

 439 TTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTG......GCTGC 482
           :::|||:::|||          |||::::::|||:::     :::::
 146 ArgAspValPheIleAsnLeuAlaAspAsnArgProPheThrProSerVa 162

 483 CGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGAC............T 520
     ::::      :::   :::   ::::::   ||||||               |
 162 lGluLysHisLeuSerLeuPheSerSerLysGluAspGlyTyrTyrIleP 179

 521 TCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCAT 570
      ||       |||:::   ::::::::::::|||            |||   |||
 179 heValValGlyValGlnAlaIleAlaLysLeuPheGlyLeuLysProHis 195

 571 GTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATAT 620
     :::   ::::::               :::|||:::::::::   ::::::||
 196 IleIleSerThrAspArgLeuThrLeuProThrGlnAspLeuValSerIl 212

 621 C...GAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCA 667
      |   :::   |||   :::   ||||||   |||:::|||       :::|
 212 eAlaHisLeuHisThrIleAspGlyProPheProSerGlySerProSerT 229

 668 CGCACATTCATTTCATCGAGCCAGTC...GGCGCGAATAAAACCGTGTGG 714
      ||||||||||   |||         :::   :::   :::   :::|||:::
 229 hrHisIleHisHisIleAlaArgIleArgAsnGluArgAspValValPhe 245

 715 ACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGG 764
     |||||::::::|||:::|||::::::||||||||:::||||||:::    ||
 246 ThrIleSerPheGlnGluValLeuSerIleGlyHisLeuPheLeuLysGl 262

 765 CCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGC......CTGCAAG 808
      |   :::       :::::::::||||||||:::|||        |||
 262 yPheValLeuGlyGlnGlnIleValAlaLeuAlaGlySerAlaLeuProP 279

 809 TCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAA 858
        ::::::   :::   |||   |||:::   |||||||:::   |||:::
 279 roSerGlnArgLysTyrLeuIleThrAlaLysGlyAlaSerPheSerAsp 295

 859 CTTACCGCCGGCGAATTGGTTGACGCGGACAAC...CGCGTGATTTCCGG 905
      |||          ::::::   ::::::|||:::       :::||||||||
 296 LeuLeuProLysAspIlePheSerSerAspGluIleThrLeuIleSerGl 312

 906 TTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGAT...TATTTGG 952
      |:::   |||:::|||   :::::::::        ::::::      ||||
 312 yAspProLeuThrGlyArgLeuCysLysLysGluGluAsnProCysLeuG 329

 953 GACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAG 1002
      ||             :::   |||:::::::::   :::   :::::::::|||
 329 lyMetArgAspHisThrIleThrLeuLeuProAsnProLysThrArgGlu 345

1003 CTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCAC 1052
          |||:::::::::                :::|||   ::::::||||||||
 346 SerPheSerPheLeuArgLeuGlyTrpAsnLysLeuThrValThrArgTh 362

1053 CACTCTCGGCCATTTCCTAAAAAACAAA......CTCTTCAAGTTCACGA 1096
      |   |||:::   |||:::|||:::|||        ::::::   :::   |
 362 rTyrLeuSerGlyPhePheLysArgLysArgValPheMetAspMetAspT 379

1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAG 1146
      ||   ::::::|||          |||   ::::::          ||||||
 379 hrAsnMetHisGlyGluLysArgProIleIleAspAlaGluIleTyrGlu 395

1147 CGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAAT 1196
     ||||||      :::   :::       :::|||:::::::::   |||
 396 ArgValSerAlaIleProValProValAlaLeuIleIleLysAlaLeuGl 412

1197 CGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACG 1246
       :::   :::   ::::::|||         ||||||   ||||||:::

 412 uThrGlnAsnPheGluGluAlaCysArgLeuGlyLeuLeuGluValAlaP 429

1247 AAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATAC 1296
        ||||||:::||||||   :::|||:::   |||:::|||   |||
 429 roGluAspPheAlaLeuProThrPheIleAspProSerLysThrGluMet 445

1297 GGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAA 1338
           ::::::::::::   |||             ||||||
 446 PheSerIleValLysGluSerLeuLeuArgTyrAlaLysGlu 459
```

```
seq_name: SwissProt_40:RNFC_VIBCH

seq_documentation_block:
ID   RNFC_VIBCH     STANDARD;      PRT;   774 AA.
AC   Q9KT88;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Electron transport complex protein rnfC.
GN   RNFC OR VC1015.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: May be part of a membrane complex involved in electron
CC       transport (By similarity).
CC   -!- COFACTOR: Binds 2 4FE-4S clusters (Potential).
CC   -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC       rnfD, rnfE and rnfG (By similarity).
CC   -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC       RNFC SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE004183; AAF94176.1; ALT_INIT.
DR   TIGR; VC1015; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001949; Complex1_51K.
DR   Pfam; PF01512; Complex1_51K; 1.
DR   Pfam; PF00037; fer4; 2.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW   Complete proteome.
FT   METAL       378    378       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       381    381       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       384    384       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       388    388       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       417    417       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       420    420       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       423    423       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       427    427       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   774 AA;  83500 MW;  44EB13EC304E3A4E CRC64;
```

```
alignment_scores:
          Quality:  172.00          Length:    472
            Ratio:  0.782             Gaps:     17
 Percent Similarity:  46.610   Percent Identity: 21.398

alignment_block:
 US-09-303-518D-131 x RNFC_VIBCH    ..

 Align seg 1/1  to: RNFC_VIBCH  from: 1  to: 774

     52 CAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGA 101
         |||   |||   ::::::          :::|||::::::|||       ::
     32 GlnProIleIleAsnAlaSerIleProAsnGluLeuValLeuProLeuLy 48

    102 AGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCG 151
        ::::::::::|||          ::::::   :::|||   |||:::   |
     48 sGlnHisIleGlyLysAlaGlyAspLeuLeuValLysValGlyAspArgV 65

    152 TCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTA 201
         ||   |||||||||   |||   :::       ::::::      :::
     65 alLeuLysGlyGlnProLeuThrGlnTyrThrSerThrPheMetLeuPro 81

    202 TTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAA 251
        :::   ||||||:::||||||   |||:::||||||          |||
     82 IleHisAlaProThrSerGlyValIleSerAlaIle.........GluPr 95

    252 GCGCGTACTTCAG......................TCAGTCGTGATTG 277
          |||:::::                              :::|||::::
     95 oArgThrValAlaHisProSerGlyLeuSerGluLeuCysIleValLeuT 112

    278 CCGTTGAAGGCAACGACGAA...................ATCGAGTTC 306
        ::    :::   ::::::|||                      :::|||
    112 hrProAspGlnGlnGluGluTrpPheGluLeuGlnProGlnProAspPhe 128

    307 GAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCG 356
        ::::::   :::||||||:::|||   :::|||       :::
    129 GlnGlnLeuThrProGluThrLeuLeuGluLeuIleArgGlnAlaGlyIl 145

    357 CCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGT 406
              ::::::       :::|||:::   ||||||
    145 eSerGlyMetGlyGlyAlaGlyPheProThrAla................ 156

    407 TCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAAT 456
           :::|||:::   :::       ::::::        :::::::::|||
    157 ..LysLysLeuGlnSerGlyLeuSerArgThrGluIleLeuIleIleAsn 172

    457 GCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGA 506
         |||   :::   :::|||        :::        |||::::::::::::
    173 AlaAlaGluCysGluProTyrIleThrAlaAspAspValLeuMetArgGl 189

    507 AGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCG 556
        :    |||:::::::::   :::|||:::   ::::::::::::::     :
    189 nTyrAlaHisGluIleIleGlnGlyIleGluIleValGluHisIleLeuL 206

    557 AACGTAAAATCCATGTG................................573
        ::    |||:::   :::
    206 ysProLysLeuThrIleIleGlyIleGluAspAsnLysProGluAlaVal 222

    574 ......TGTAAAGCAGCAGGCGCAGACGTGCCGTCT.............. 603
                 :::    |||:::    |||    |||
    223 AlaAlaLeuGlnGlnAlaAlaGlnAspLysProMetValIleArgValIl 239

    604 .......................................GAAAATGCTG 613
                                                :::     :
    239 eProThrLysTyrProSerGlyGlyGluLysGlnLeuIleLysIleLeuT 256

    614 CCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGT 663
        ::|||:::|||:::   :::    |||    |||         ||||||
    256 hrAsnLeuGluValProLysGlyGlyIleProAlaAspIleGlyLeu... 271

    664 GGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTG 713
                       ::::::::   |||||||:::
    272 ..............MetValGlnAsnValGlySer............. 278

    714 GACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAG 763
                          :::    ||||||:::|||    :::|||    |
    279 ..................LeuGlnAlaIleAlaArgAlaIleValHisG 289

    764 GCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAAC 813
        ||:::           :::|||||||||||:::|||    |||          :::
    289 lyGluProLeuIleArgArgValValThrLeuThrGlyAspCysPheArg 305

    814 AAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTAC 863
        |||||||||   :::    ::::::|||||||:::     |||:::    |||
    306 LysProArgAsnValTrpAlaLeuLeuGlyThrProValGlnAlaLeuLe 322

    864 CGCCGGCGAATTGGTTGACGCGGACAAC.........CGCGTGATTTCCG 904
                          ||||||:::          |||:::|||    |
    322 uAsnGluPheGlyTyrLysAlaAspLysLysLeuProArgLeuIleMetG 339

    905 GTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGA 954
        ||:::    :::    |||
    339 lyGlyProMetMetGlyPheThrLeuPro..................... 348

    955 CGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCT 1004
              |||    |||:::    :::                 :::|||
    349 ......HisAlaGlnValProIle..............ThrLysThrAl 358

   1005 GTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCA 1054
                   :::||||||                      ||||||:::
    358 aAsnCysIleLeuAlaPro....................ThrArgAsnG 368

   1055 CTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTC 1104
           |||                                  |||::::::
    368 luLeu..............................ThrSerSerAsp 373

   1105 AACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAAT 1154
        |||        :::    |||    :::          |||          ::::::
    374 Asn......GluMetAlaCysIleArgCysGlyGlnCysAlaGluAlaCy 388

   1155 GCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCG 1204
          |||::::::::||||||       |||                 :::   :
    388 sProValSerLeuLeuProGlnGlnLeuGlnTrpHisAlaLysAlaGluG 405

   1205 ATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGAC 1254
        ::    |||:::                 |||    |||||||||    :::|||
    405 luPheAspLys..............CysGluGluLeuAspLeuLysAsp 416

   1255 .........CTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATA 1295
                     :::    |||::::::|||||||||||::::::
    417 CysIleGluCysGlyAlaCysAlaTyrValCysProSerGluIle..... 431

   1296 CGGCCCGCTGTTGCGC 1311
            ||||||::::::
    432 ....ProLeuValGln 435

seq_name: SwissProt_40:YM96_YEAST

seq_documentation_block:
ID   YM96_YEAST     STANDARD;      PRT;  1140 AA.
AC   Q04893;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.
GN   YMR317W OR YM9924.09.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z54141; CAA90835.1; -.
DR   SGD; S0004936; YMR317W.
KW   Hypothetical protein; Repeat.
SQ   SEQUENCE   1140 AA;  113070 MW;  0153EBCA24FE5427 CRC64;

alignment_scores:
             Quality:  143.00          Length:    480
               Ratio:  0.633             Gaps:     17
 Percent Similarity:  47.083   Percent Identity: 22.292

alignment_block:
 US-09-303-518D-131/rev x YM96_YEAST   ..

 Align seg 1/1  to: YM96_YEAST  from: 1  to: 1140

    1331 TCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTCGTATTTGCC 1282
          |||   :::||||||:::::   :::::::::|||   |||
     386 SerSerThrSerSerSerValSerSerGluAlaProSerSer........ 399

    1281 CGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGC 1232
                                 |||||||||   :::|||:::
     400 ........................ThrSerSerSerValSerSerGluA 408

    1231 AACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCGCGCAAA 1182
            |||:::::   :::   |||||||||      |||:::|||   |||
     408 laProSerSerThrSerSerSerValSerSerGluIleSerSerThrLys 424

    1181 AGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCGAT 1132
          |||:::|||      |||||||:::   ::::::   :::
     425 SerSerVal......MetSerSerGluValSerSerAla.........Th 436

    1131 CGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTGAAGA 1082
           ::::::::::::   |||        :::|||:::   :::|||    |
     436 rSerSerLeuValSerSerGluAlaProSerAlaIleSerSerLeuAlaS 453

    1081 GT......TTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATG... 1041
          ||       ||||||      |||         |||:::   ::::::
     453 erSerArgLeuPheSerSerLysAsnThrSerValThrSerThrLeuVal 469

    1040 ......GAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTT 997
                |||       |||      ::::::   :::|||:::|||:::
     470 AlaThrGluAlaSerSerValThrSerSerLeuArgProSerSerGluTh 486

     996 GCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAAT 947
           |||         :::||||||   |||
     486 rLeuAlaSerAsnSerIleIleGlu......................... 494

     946 AATCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATC 897

     494 .................................................. 494

     896 ACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTT 847
          :::     ||||||::::::   ||||||   :::   :::      :::
     495 SerSerLeuSerThrGlyTyrAsnSerThrValSerThrThrThrSerAl 511

     846 CGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGC 797
            |||   :::|||:::   ::::::   :::      :::   |||
     511 aAlaSerSerThrLeuGlySerLysValSerSerSerAsnSerArgMetA 528

     796 CCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGT 747
               :::   |||   |||
     528 laThrSerLysThrSerSer.............................. 534

     746 CCGATAGCAATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGC 697
                        ||||||      |||         :::|||:::|||::
     535 ...........ThrSerSerAspLeuSerLysSerSerValIlePheGl 547

     696 GCCGACTGGCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGAT 647
          :   :::::::::::
     547 yAsnSerSerThrVal.................................T 553

     646 GCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAGCA............ 609
                 |||:::::   :::|||:::   |||:::
     553 hrThrSerProSerAlaSerIleSerLeuThrAlaSerProLeuProSer 569

     608 ...TTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTT 562
            :::||||||   ||||||:::   |||:::      :::      ||
     570 ValTrpSerAspIleThrSerSerGluAlaSerSerIleSerSerAsnLe 586

     561 ACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGG 512
          |   |||            :::   |||:::      :::   |||
     586 uAlaSerSerSerAlaProSerAspAsnAsnSerThrIleAlaSer.... 601

     511 CGGCTTCTTTGATGATGACC..................GTAGGGTCGGCA 480
            |||||||||::::::|||                    |||:::|||
     602 ..AlaSerLeuIleValThrLysThrLysAsnSerValValSerSerIle 617

     479 GCCAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGC 430
          :::|||:::::::::|||         |||:::   :::|||   :::::
     618 ValSerSerIleThrSerSerGluThrThrAsnGluSerAsnLeuAlaTh 634

     429 ATCTACGGCAGGGATTTTGCTGAAC......................... 405
          :|||:::::::::::|||   |||
     634 rSerSerThrSerLeuLeuSerAsnLysAlaThrAlaArgSerLeuSerT 651

     404 ..GGACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGG 357
           :::::::::   ||||||   :::   |||
     651 hrSerAsnAlaThrSerAlaSerAsnValPro..............Thr 662

     356 CGCACTTTTTCGCTGCTCAAT..........................TT 334
             |||||||||   ::::::                           ||
     663 GlyThrPheSerSerMetSerSerHisThrSerValIleThrProGlyPh 679

     333 TGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATT.........TCGT 293
          |:::::::::|||::::::         ||||||          ||||
     679 eSerThrSerSerAlaSerLeuAlaIleAsnSerThrValValSerSerS 696

     292 CGTTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCA 243
          |||||   :::   :::::::::|||      |||:::      :::
     696 erLeuAlaGlyTyrSerPheSerThrProGluSerSerProThrThrSer 712

     242 CGGTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCC 193
               :::::::::         |||::::::   :::      ||||||
     713 ThrLeuValThrSerGluAlaProSerThrValSerSerMetThrThrSe 729

     192 CGGATTCTTTTTGTCT.....................TCAAACAGCACTT 164
           :::   |||::::::                       |||   ||||||
     729 rAlaProPheIleAsnAsnSerThrSerAlaArgProSerProSerThrA 746

     163 GGCCTTTTTTGACGGCTTCACCTTCCTTGATT.................. 132
             |||:::|||   |||   |||   |||
     746 laSerPheIleThrGluSerThrSerSerIleSerSerValProLeuAla 762

     131 .....................TTCATCGAGGGGCGCATGCCGACATATTC 103
                              :::         ::::::   |||:::|||
     763 SerGlyAspValThrSerSerLeuAlaAlaHisAsnLeuThrThrPheSe 779

     102 TTCGCCAAGCAACGCGACTTCGGTAATGGCCGGGCCGTCA 63
          |:::|||||||:::::::::::   :::::::::   |||
     779 rAlaProSerThrSerSerAlaGlnLeuValSerLysSer 792

seq_name: SwissProt_40:RNFC_HAEIN

seq_documentation_block:
ID   RNFC_HAEIN     STANDARD;      PRT;   819 AA.
```

```
AC   P71397;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Electron transport complex protein rnfC.
GN   RNFC OR HI1685.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: May be part of a membrane complex involved in electron
CC       transport (By similarity).
CC   -!- COFACTOR: Binds 2 4FE-4S clusters (Potential).
CC   -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC       rnfD, rnfE and rnfG (By similarity).
CC   -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC       RNFC SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32841; AAC23331.1; -.
DR   HSSP; P00198; 2FDN.
DR   TIGR; HI1685; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001949; Complex1_51K.
DR   Pfam; PF01512; Complex1_51K; 1.
DR   Pfam; PF00037; fer4; 2.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW   Complete proteome.
FT   METAL       378    378       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       381    381       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       384    384       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       388    388       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       417    417       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       420    420       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       423    423       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       427    427       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   819 AA;  89283 MW;  541EAA2AEC6AEA93 CRC64;

alignment_scores:
          Quality:  135.50           Length:    465
            Ratio:  0.622              Gaps:     15
 Percent Similarity:  46.882   Percent Identity: 19.570

alignment_block:
 US-09-303-518D-131 x RNFC_HAEIN   ..

 Align seg 1/1  to: RNFC_HAEIN  from: 1  to: 819
```

```
 70 CCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCG 119
    |||     |||:::       :::      ::::::::::|||
 38 ProLeuGlyThrAspPheTyrIleProLeuLysGlnHisLeuGlyThrTh 54

120 CCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGC 169
      ::::::  ||||||||||||:::  |||  |||||||||:::|
 54 rGlyAsnLeuLeuIleLysGluGlyAspTyrValLeuLysGlyGlnAlaL 71

170 TGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCA 219
    ||   :::                 :::          ||||||:::|||
 71 euThrLysGlyAspGlyLeuArgMetLeuProValHisAlaProThrSer 87

220 GGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGT 269
    |||   |||                           |||  :::     ||
 88 GlyThrIle....................LysSerIleLysProTyrVa 97

270 CGTGATTGCCGTTGAAGGCAACGACGAAATCGAG................ 303
    |:::           :::|||  ||||||
 97 lAlaThrHisProSerGlyLeuAspGluProThrIleHisLeuGlnAlaA 114

304 .................TTCGAACGCTACGTACCTGAAGCGCTGGCAAAA 336
                    :::||||||        :::   ::::::
114 spGlyLeuAspGlnTrpIleGluArgAsnProIleAspAspPheSerThr 130

337 TTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGAC 386
    ||||||||||||::::::   :::::::::   |||:::|||:::    ::
131 LeuSerSerGluGlnLeuIleHisLysIleTyrGlnAlaGlyIle...Al 146

387 TGCGCTTCGCACCCGTCCGTTC.........AGCAAAATCCCTGCCGTAG 427
    ::::|||                 |||       :::||||||  ::::::
146 aGlyLeuGlyGlyAlaValPheProThrAlaAlaLysIleGlnSerAlaG 163

428 ATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTG 477
    ::   :::        :::::::::|||:::   :::   :::|||
163 luGlnLysValLysLeuLeuIleIleAsnGlyAlaGluCysGluProTyr 179

478 GCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACG 527
       :::          :::::::::|||   |||:::::::::   ::
180 IleThrCysAspAspArgLeuMetArgGluArgAlaAspGluIleIleLy 196

528 CGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTA 577
    :|||:::   :::|||        :::     :::     :::     |||
196 sGlyIleArgIleLeuArgTyrIleLeuHisProGlu...LysValValI 212

578 AAGCAGCAGGCGCAGACGTGCCGTCT........................ 603
      |||           :::   |||:::
212 leAlaIleGluAspAsnLysProGluAlaIleSerAlaIleArgAsnAla 228

604 ...GAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
      ::::::|||   :::|||:::   :::     :::           :::||
229 LeuGlnGlyAlaAsnAspIleSerIleArgValIleProThrLysTyrPr 245

651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAG............C 688
    |:::|||   :::             |||:::::::::            |
245 oSerGlyAlaThrLysGlnLeuIleTyrLeuLeuThrGlyIleGluValP 262

689 CAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACGTG... 735
    ||     |||    :::::::::          :::        |||:::|||
262 roSerGlyGluArgSerSerSerIleGlyValLeuMetGlnAsnValGly 278

736 ......ATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGA 779
          :::||||||   |||   :::::::::   :::            ||
279 ThrMetPheAlaIleLysArgAlaIleIleAsnAspGluProLeuIleGl 295

780 GCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGC 829
    |||||||||:::|||   |||   ::::::   :::
295 uArgValValThrLeuThrGlyAsnLysIleAlaGluLysGlyAsnTyrT 312

830 GTACCGTTTTGGGTGCGAAGGTGTCTCAACTT......ACCGCCGGCGAA 873
       :::   |||||||:::   :::||||||:::            ||||||
312 rpValArgLeuGlyThrProIleSerGlnIleLeuSerAspAlaGlyTyr 328
```

```
 874 TTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGC 923
           |||             :::::::::|||:::    :::    |||
 329 GlnPheAspLysHisPheProIlePheAlaGlyGlyProMetMetGlyLe 345

 924 GATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGATTT 973
                                             ::::::
 345 u.........................................GluLeuP 348

 974 CCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCG 1023
           ::::::        :::|||   :::         :::||||||
 348 roAsnLeuAsnAlaProValThrLysLeuValAsnCysLeuLeuAlaPro 364

1024 CAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAA 1073
      :::       :::|||
 365 AspTyrLeuGluTyr.................................. 369

1074 AAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCA 1123
                                 |||      :::::::::|||
 370 .........................AlaGluProGluAlaGluGlnAlaC 378

1124 TGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCT 1173
        :::      ::::::   :::   :::    |||::::::::::::|||
 378 ysIleArgCysSerSerCysSerAspAlaCysProValAsnLeuMetPro 394

1174 ACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGC 1223
           |||                     |||      :::::::::
 395 GlnGlnLeuTyrTrpPheAlaArgSerGluAspHisLysLysSerGluGl 411

1224 TTTGGGT.........TGCTTGGAATTGGACGAAGAAGACCTCGCTTTGT 1264
        :::        |||:::|||                 ::::::|
 411 uTyrAlaLeuLysAspCysIleGlu..............CysGlyIleC 423

1265 GCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGC... 1311
     ||::::::|||||||||          ::::::   ||||||:::::
 423 ysAlaTyrValCysPro.........SerHisIleProLeuIleGlnTyr 436

1312 ..................AAAGTGCTGGAAACCATTGAGAAGGAA 1338
                        |||:::    :::    ||||||:::
 437 PheArgGlnGluLysAlaLysIleTrpGlnIleLysGluLysGln 451
```

seq_name: SwissProt_40:RNFC_PSEAE

seq_documentation_block:

```
ID   RNFC_PSEAE     STANDARD;      PRT;   774 AA.
AC   Q9HYB8;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Electron transport complex protein rnfC.
GN   RNFC OR PA3491.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: May be part of a membrane complex involved in electron
CC       transport (By similarity).
CC   -!- COFACTOR: Binds 2 4FE-4S clusters (Potential).
CC   -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC       rnfD, rnfE and rnfG (By similarity).
CC   -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC       RNFC SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE004770; AAG06879.1; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001949; Complex1_51K.
DR   Pfam; PF01512; Complex1_51K; 1.
DR   Pfam; PF00037; fer4; 2.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW   Complete proteome.
FT   METAL       369    369       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       372    372       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       375    375       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       379    379       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       408    408       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       411    411       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       414    414       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       418    418       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   774 AA;  82359 MW;  C82B06F4D2E09BB6 CRC64;
```

alignment_scores:

```
           Quality:  132.00             Length:    492
             Ratio:  0.600                Gaps:     24
 Percent Similarity:  44.715  Percent Identity:  20.935
```

alignment_block:
US-09-303-518D-131 x RNFC_PSEAE ..

Align seg 1/1 to: RNFC_PSEAE from: 1 to: 774

```
  28 CTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCAT 77
     |||||||||         |||          |||      |||
  23 LeuProIleGlnGlnAlaProLeuAlaGlnArgTyrIleValPro..... 37

  78 TACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGC......ATGCGCC 121
                       ||||||   :::::::::|||          ||||
  38 ................LeuGly...GlnHisIleGlyAlaProAlaArgP 48

 122 CCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTG 171
     ||    :::::::::        |||:::||||||   |||||||||::::::
  48 roCysValGluVal......GlyGlnAlaValLeuLysGlyGlnThrIle 62

 172 TTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGG 221
                      :::   :::::::::::    ||||||:::|||||
  63 AlaLeuProAspGlyThrValSerAlaAlaLeuHisAlaProThrSerGl 79

 222 CAAAATCGCCGCTATT..............CACCGTGGCGAAAAGCGCG 256
     |    ::::::||||||                    :::::::::
  79 yThrValValAlaIleGlyAlaHisProTyrProHisAlaSerGlyLeuP 96

 257 TACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAA......... 297
              ::::::|||||||||   :::|||    ::::::
  96 roAlaProAlaIleValIleAlaSerAspGlyLeuGluArgTrpThrGlu 112

 298 ...........ATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAA 335
                    :::|||:::          |||   ||||||
 113 LeuHisProCysProAspPheArgAlaGluSerProLeuAlaLeuLeu.. 128

 336 ATTGAGCAGCGAAAAAGTGCGCCGC...........AACCTGATTCAAT 373
               |||:::::::|||                :::|||      :
 129 .........GluArgIleArgAlaAlaGlyIleGlyGlyLeuGlyGlyA 142

 374 CAGGCTTATGGACTGCG......CTTCGCACCCGTCCGTTCAGCAAAATC 417
```

```
         ::|||:::   |||||     |||   :::||||||   :::||||||
 142 laGlyPheProThrAlaAlaLysLeuAlaAlaArgProAlaGluLysIle 158

 418 CCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACAC 467
                                :::::: |||||||:::   :::
 159 ....................HisThrLeuValValAsnGlyAlaGluCy 168

 468 CAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAG 517
      :::|||     :::            ::::::::::|||   |||   :
 168 sGluProTyrIleSerAlaAspAspLeuLeuMetArgGluArgAlaThrG 185

 518 ACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATC 567
     ::          |||:::   :::|||    ::::::
 185 lnValLeuGlyGlyIleAspIleLeuValGlnIle.............. 196

 568 CATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAA 617
        :::|||                   |||:::|||     :::::
 197 ...LeuCys....................ProGluGluValLeuValGl 205

 618 TATCGAAACACATGAA.................TTTGGCGGCCCG.... 645
     :||||||    :::                   :::|||:::
 205 yIleGluAspAspLysProGluAlaIleAlaAlaLeuGlyAlaAlaLeuG 222

 646 ....................................CATCCTGCCGGC 657
                                         :::|||:::|||
 222 lyGluArgProTyrArgIleValAlaLeuProThrArgTyrProSerGly 238

 658 .......................TTGAGTGGCACGCACATTCATTTCAT 683
                            |||:::|||   ::::::
 239 GlyGluArgGlnLeuIleGlnLeuLeuThrGlyArgGluValPro...Al 254

 684 CGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACG 733
      :::   :::   |||:::      :::         |||         :
 254 aAspGlyLeuProAlaAspIleGlyIleLeuCysGlnAsnValGlyThrL 271

 734 TGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGC 783
     ::   |||:::             |||   ||||||        :::|||
 271 euAlaAlaValHisAspAlaValValLeuGlyArgProLeuIleSerArg 287

 784 GTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTAC 833
     :::::::::|||:::|||        :::::::::|||      ::::::::
 288 IleThrThrLeuAlaGlyGlyAlaLeuGluArgProMetAsnValGluAl 304

 834 CGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCC................ 867
     :::::::|||:::   |||   :::|||   |||
 304 aLeuIleGlyThrProValHisGluLeuLeuAlaPheAlaGlyLeuAlaG 321

 868 ..GGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTG 915
        |||:::|||                :::|||||||:::   |||:::
 321 luGlyArgLeu............GluArgValLeuMetGlyGly...... 331

 916 AACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAA 965

 331 .................................................. 331

 966 TCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGG 1015
                                        ::::::|||::::
 332 ................................ProMetMetGlyPheA 337

1016 TTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCAT 1065
     ::    |||:::
 337 laLeuProAsp...................................... 340

1066 TTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGA 1115
        |||:::      |||:::|||      :::::::::   :::||||||
 341 ...LeuSerValProLeuIleLysThrCysAsnCysLeuLeuAlaGlyAs 356

1116 CCGCGCCATGGTACCG.............................ATCG 1135
     |    :::    :::|||                              |
 356 pAlaThrGluLeuProGluProValProAlaMetProCysIleArgCysG 373

1136 GCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTG 1185
      ||          :::|||   |||:::::::::||||||        |||
 373 lyAspCysAlaGlnValCysProValSerLeuLeuProGln...GlnLeu 388

1186 CGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTT 1235
      :::   :::   :::||||||    :::        |||   |||:::
 389 HisPhePheAlaLeuGlyAsp...GluHisGluGlnLeuLeuAlaHisAs 404

1236 GGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCA 1285
          :::|||         :::   :::   |||::::::|||||||||::::
 404 nLeuPheAspCysIleGluCysGlyAlaCysAlaTyrValCysProSerS 421

1286 AATACGAATACGGCCCGCTGTTGCGC 1311
      ::               ||||||::::::
 421 erIle.........ProLeuValGln 426
```

seq_name: SwissProt_40:RNFC_BUCAI

seq_documentation_block:

```
ID   RNFC_BUCAI     STANDARD;      PRT;   473 AA.
AC   P57215;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Electron transport complex protein rnfC.
GN   RNFC OR BU115.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TOKYO 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- FUNCTION: May be part of a membrane complex involved in electron
CC       transport (By similarity).
CC   -!- COFACTOR: Binds 2 4FE-4S clusters (Potential).
CC   -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC       rnfD, rnfE and rnfG (By similarity).
CC   -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC       RNFC SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AP001118; BAB12833.1; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001949; Complex1_51K.
DR   Pfam; PF01512; Complex1_51K; 1.
DR   Pfam; PF00037; fer4; 2.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Electron transport; Iron-sulfur; 4Fe-4S; Complete proteome.
FT   METAL       337    337       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       340    340       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       343    343       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       347    347       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       376    376       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       379    379       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       382    382       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       386    386       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   473 AA;  54563 MW;  04E64102F1315AEE CRC64;
```

alignment_scores:
Quality: 131.00   Length: 440

```
              Ratio:  0.618              Gaps:     20
 Percent Similarity:  48.182   Percent Identity:  20.455

alignment_block:
 US-09-303-518D-131 x RNFC_BUCAI   ..

 Align seg 1/1   to: RNFC_BUCAI  from: 1   to: 473

     127 ATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAA...GTGCTGTT 173
          ::::::::|||   :::|||   |||   :::||||||   ::::::||
      20 LeuArgValLysIleAsnGluLysValLeuArgGlyGlnProLeuIlePh 36

     174 TGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCA 223
          |:::|||      |||      |||      . ||||||:::||||||
      36 eSerAspAspPheAsn......ValProValHisAlaProThrSerGlyL 51

     224 AAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTG 273
            |||      |||      ::::::    :::   :::::::::
      51 euIleGluAsnIleCysPheAsnSerAspSerIleLysLysAsnIleLys 67

     274 ATTGCCGTTGAAGGC.........................AACGACGA 296
          |||:::::::::                            |||
      68 IleValIleSerProAspTyrLeuAspGlnTrpIleArgLeuAsnProIl 84

     297 AATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCG 346
             :::::::::::|||:::||||||    |||    |||:::
      84 eLysAspTyrLysLysTyrAlaProGluLysLeuIleLysIle...... 98

     347 AAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGC 396
                            :::   |||||||||:::    :::::::|||
      99 .................IleHisGlnSerGlyVal...ValGlyLeuGly 108

     397 ACCCGTCCGTTC.........AGCAAAATCCCTGCCGTAGATGCCGAGCC 437
                |||            :::||||||            :::
     109 GlyGlyGlnPheProSerSerLysLysIleIlePheSerIleAsnArgAl 125

     438 GTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACC 487
             :::::::::|||||||||::::::::::::|||      ::::::
     125 aHisThrLeuIleValAsnAlaValGluSerGluProTyrIleThrSerA 142

     488 CTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTG 537
            :::   :::|||   :::       :::::::::      |||
     142 spAsnCysLeuIleTyrAsnHisIleSerGluIleLeuIleGlyCysLys 158

     538 GTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGG 587
          ::::::      :::|||:::   |||       :::
     159 IleIleCysTrpIleThrLysIleLysThrValLeuIleAlaIleGlnGl 175

     588 CGCAGACGTGCCGTCT.......................GAAAATGCTG 613
             ::::::   |||                          :::|||   :
     175 uAspAsnIleGlnSerIleSerLysIleGlnHisLeuIleLysAsnLysS 192

     614 CCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGC...... 657
          ::   ::::::         :::         :::|||||||||
     192 erLeuPheLysIleCysIleIleLysLysLysTyrProAlaGlySerSer 208

     658 .................TTGAGTGGCACGCACATT...........CA 677
                           |||:::|||   ::::::           ||
     209 LysValLeuValLysSerLeuThrGlyLysGluValProHisGlyLysHi 225

     678 TTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATC 727
          |   |||:::   :::|||                  :::|||
     225 sSerIleAsp...IleGly...........TyrLeuIlePheAsnValA 237

     728 AAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACC 777
             :::::::::|||   |||   ::::::::::|||:::       |||
     237 laThrIlePheSerIleLysArgAlaIleIleAsnGlyLysProLeuThr 253

     778 GAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTT 827
          ||||||||||||:::|||                ::::::      :::|||||
     254 GluArgValValThrLeu...........MetSerAspLysAsnLeuLe 266

     828 GCGTACCGTTTTGGGTGCGAAGGTGTCTCAA..............CTTA 862

         |         :::   :::::::::::                  ||||
     266 uSerGlyAsnPheTrpValArgIleGlyThrProIleLysTyrPheLeuT 283

     863 CCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTA 912
          ||::::::::|||     ::::::              |||::::::
     283 hrSerAsnLysLeuLysGlnSer.........PheIleAlaSerValTyr 296

     913 TTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCA 962
          |||:::|||                           :::::|||:::
     297 LeuGlyGly...................ProPheMetGlyLysLysIl 306

     963 CAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCT 1012
            |||:::::::::
     306 eAsnAsnLeuAsn..................................... 310

    1013 GGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGC 1062
                                  :::||||||   :::    |||    :::
     311 .....................HisSerIleLeuLysLysThrAsnSer 319

    1063 CATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGG 1112
             |||:::    :::|||       |||       :::            ::
     320 IlePheIleThrHisLysLysGluLysAsnGluSerIle.........Se 333

    1113 CGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGG 1162
          ::::::::::    :::        |||       :::    |||    |||:::
     333 rGluLysThrCysIleArgCysGlyTyrCysSerTyrValCysProValA 350

    1163 ACATCCTGCCTACCTTGCTT.........TTGCGCGATTTAATCGTCGGC 1203
          :::::||||||        |||             :::::::::
     350 snLeuLeuProGlnGlnLeuTyrTrpTyrIleLysAsnLysAsnHisVal 366

    1204 GATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGA 1253
          :::|||    :::        :::|||    |||:::|||
     367 GlnThrLysLysHisTyrValLeuAspCysIleGlu............. 378

    1254 CCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGC 1303
                       |||:::    |||||||||          :::|||    ||||
     379 .CysLysAlaCysGluLysValCysPro.........SerTyrIleProL 392

    1304 TGTTGCGCAAAGTGCTGGAA 1323
          ||::::::       ::::::
     392 euValLysTyrPheIleGln 398

seq_name: SwissProt_40:N121_RAT

seq_documentation_block:
ID   N121_RAT       STANDARD;      PRT;  1199 AA.
AC   P52591;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE   of 121 kDa) (P145).
GN   NUP121.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX   MEDLINE=93328754; PubMed=8335683;
RA   Hallberg E., Wozniak R.W., Blobel G.;
RT   "An integral membrane protein of the pore membrane domain of the
RT   nuclear envelope contains a nucleoporin-like region.";
RL   J. Cell Biol. 122:513-521(1993).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE
CC       REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS
CC       OF THE PORE COMPLEX TO THE PORE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC       MEMBRANE.
CC   -!- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC   -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
```

```
CC       SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC       NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z21513; CAA79725.1; -.
DR   EMBL; Z21514; CAA79726.1; -.
KW   Nuclear protein; Transport; Transmembrane; Repeat.
FT   DOMAIN        1     56       CISTERNAL SIDE (POTENTIAL).
FT   TRANSMEM     57     77       POTENTIAL.
FT   DOMAIN       76   1199       PORE SIDE (POTENTIAL).
FT   DOMAIN        4      8       POLY-ALA.
FT   DOMAIN       53     58       POLY-ALA.
FT   DOMAIN      438    441       POLY-SER.
FT   DOMAIN      497    500       POLY-PRO.
FT   DOMAIN      684    687       POLY-SER.
FT   DOMAIN     1023   1026       POLY-ALA.
FT   DOMAIN     1033   1038       POLY-SER.
FT   DOMAIN     1058   1063       POLY-GLY.
SQ   SEQUENCE   1199 AA;  120784 MW;  6DC4451B91D5B907 CRC64;

alignment_scores:
          Quality:  129.50            Length:    478
            Ratio:  0.576               Gaps:     20
 Percent Similarity:  47.071    Percent Identity:  22.385

alignment_block:
 US-09-303-518D-131/rev x N121_RAT    ..

 Align seg 1/1  to: N121_RAT  from: 1  to: 1199

    1340 CCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTC 1291
           |||  |||:::::::::       |||:::     ::::::    ::
     546 ProProPheThrLeuThrLeuProThrValGlyProAlaAlaSerProAl 562

    1290 GTATTTGCCCGGGCAGACG.................AAGCTGCACAAAG 1259
           :  ||||||:::   :::                  :::|||   |||
     562 aSerLeuProAlaProSerSerAsnProLeuLeuGluSerLeuLysLysM 579

    1258 CGAGGTCTTCT...TCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTG 1212
            ::::::|||   :::   :::||::::::|||      :::|||
     579 etGlnGluSerProAlaProSerSerSerGluProProGluAlaAlaThr 595

    1211 TCGGTATCGCCGACGATTAAATCGCGCAAAAGCAAGGTAGGCAGGATGTC 1162
            ::::::|||:::       :::        |||    ::::::   :::
     596 ValAlaAlaProSerProProLysThrProSerLeuLeuAlaProLeuVa 612

    1161 CAACGGCATTACGCGCTCATAAGTGCCGATC...GGTACCATGGCGCGGT 1115
           :::   :::|||                |||:::   :::|||   :::   |
     612 lSerProLeuThrGly.........ProLeuAlaSerThrSerSerAspS 626

    1114 CGCCGCCGTTGACGGCTGTCGTGAACTTGAAGAGT.............. 1080
          ||   |||   |||:::   ::::::|||   |||
     626 erLysProThrThrThrPheLeuGlyLeuAlaSerAlaSerSerAlaThr 642

    1079 ...TTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTT 1033
            |||          |||   |||   |||   :::::::::       ::
     643 ProLeuThrAspThrLysAlaProGlyValSerGlnAlaGlnLeuCysVa 659

    1032 GTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCT............... 999
          :|||       :::||||||   |||:::   :::
     659 lSerThrProAlaAlaThrAlaProSerProThrProAlaSerThrLeuP 676

     998 ........TTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAG 957
                 |||       |||:::|||   :::::::::
     676 heGlyMetLeuSerProProAlaSerSerSerSerLeu.........Ala 689

     956 CGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCGTTC........ 915
           |||       :::||||||     :::    |||:::
     690 ThrProGlyProAlaCysAlaSerProMetPheLysProIlePheProAl 706

     914 ................AATACCGAACCGGAAATCACGCGGTTGTCCGCGT 881
                         :::   :::|||       |||       ||||||:
     706 aThrProLysSerGluSerAspAsnProLeuProThrSerSerSerAlaA 723

     880 CAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAAAACGGTA 831
          ::|||::::::||||||
     723 laThrThrThrProAla................................ 728

     830 CGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAACC...AC 784
                             |||:::   |||          ||||||   ||
     729 ..................SerThrAlaLeuProThrThrAlaThrAlaTh 739

     783 GCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCAATCA 734
          |      :::|||:::                  :::    :::|||:::|||
     739 rAlaHisThrPheLysProIlePheGluSerValGluProPheAla.... 754

     733 CGTCTTGATAATTGATGGTCCACACGGTTTATTCGCGCCGACTGGCTCG 684
                                ::::::    ::::::|||      |||
     755 .......................AlaMetProLeuSerProProPheSer 763

     683 ATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAA 634
          :::|||        :::|||         |||:::    :::    |||
     764 LeuLysGlnThrThrAlaProAlaThrThrAlaAlaThrSerAlaPro.. 779

     633 TTCATGTGTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTG 584
                      :::|||:::::::::::::    :::|||||||:::     |
     780 .............LeuLeuThrGlyLeuGlyThrAlaThrSerThrValA 792

     583 CTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAAC 534
          ||:::       |||                           :::::::::
     792 laThrGlyThrThr.......................AlaSerAlaSer 800

     533 AGGCCGCGTTTG.........AAGTCTTCGGCGGCTTCTTTGATGATGAC 493
          :::|||    :::               ::::::|||||||||      ||
     801 LysProValPheGlyPheGlyValThrThrAlaAlaSerThrAlaSerTh 817

     492 CGTAGGGTCGGCAGCCC...AGCGGATTGGTGTCCATCGCATTGACGAAGA 446
          |::::::|||::::::   |||    |||    :::    |||
     817 rIleAlaSerThrSerGlnSerIleLeuPheGlyGlyAlaProProValT 834

     445 TGGCGAACGGCTCGGCA...TCTACGGCAGGGATTTGCTGAACGGACGG 399
           |||::::::||||||    :::    |||:::|||:::          |||
     834 hrAlaSerSerSerAlaProAlaLeuAlaSerIlePheGlnPheGly... 849

     398 GTGCGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTT 349
                           ||||||
     850 ..................LysPro......................... 851

     348 TTCGCTGCTCAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGA 299
              |||          |||||::::::||||||          |||:::|||:
     852 ....LeuAlaProAlaAlaSerValAlaGlyThrSerPheSerGlnSerL 867

     298 TTTCGTCGTTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTT 249
          ::::::|||       :::||||||
     867 euAlaSerSerAlaGlnThrAla.......................... 874

     248 TCGCCACGGTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATAC 199
                                                 |||    |||::
     875 ....................................AlaSerAsnSe 878

     198 TACGCCCGGATTC..............TTTTTGTCTTCAAACAGCACTT 164
          ::::    ||||||                  |||:::::::::::::::
     878 rSerGlyGlyPheSerGlyPheGlyGlyThrLeuThrThrSerThrSerA 895

     163 GGCCTTTTTTGACGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGCATG 114
            |||          |||::::::|||:::|||    |||    :::
     895 laProAlaThrThrSerGlnProThrLeuThrPheSerAsnThrValThr 911
```

```
 113 CCGACATATTCTTCGCCAAGCAACGCGACTTCGGTAATGGCCGGGCCGTC 64
      ||||||::::::   |||   :::||||::::::       |||   |||::
 912 ProThrPheAsnIleProPheSerAlaSerAlaLysProAlaLeuProTh 928

  63 ATAAATGACTTGCTCCGGTCTGCCCGCGATGGGC 30
      :          :::::::::   |||::::::|||
 928 rTyrProGlyAlaAsnSerGlnProThrPheGly 939
```

seq_name: SwissProt_40:HKR1_YEAST

```
seq_documentation_block:
ID   HKR1_YEAST     STANDARD;      PRT;  1802 AA.
AC   P41809;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hansenula MRAKII killer toxin-resistant protein 1 precursor.
GN   HKR1 OR YDR420W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YNN 295;
RX   MEDLINE=94156857; PubMed=8113191;
RA   Kasahara S., Yamada H., Mio T., Shiratori Y., Miyamoto C.,
RA   Yabe T., Nakajima T., Ichishima E., Furuichi Y.;
RT   "Cloning of the Saccharomyces cerevisiae gene whose overexpression
RT   overcomes the effects of HM-1 killer toxin, which inhibits
RT   beta-glucan synthesis.";
RL   J. Bacteriol. 176:1488-1499(1994).
CC   -!- FUNCTION: COULD REGULATE BETA-GLUCAN SYNTHESIS. OVEREXPRESSION
CC       PROVIDES RESISTANCE TO HM-1 KILLER TOXIN.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- PTM: COULD BE O-GLYCOSYLATED IN SERINE/THREONINE RICH DOMAIN.
CC   -!- SIMILARITY: SOME, TO YEAST MSB2.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; S69101; AAB30051.1; -.
DR   SGD; S0002828; HKR1.
KW   Glycoprotein; Transmembrane; Repeat; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1802       HANSENULA MRAKII KILLER TOXIN-RESISTANT
FT                                PROTEIN 1.
FT   TRANSMEM   1486   1506       POTENTIAL.
FT   DOMAIN       23   1478       SER/THR-RICH.
FT   DOMAIN      453    788       12 X 28 AA TANDEM REPEATS OF S-[AV]-[P]-
FT                                V-A-V-S-S-T-Y-T-S-S-P-S-A-P-A-A-I-S-S-T-
FT                                Y-T-S-S-P.
FT   REPEAT      453    480       1 (APPROXIMATE).
FT   REPEAT      481    508       2.
FT   REPEAT      509    536       3.
FT   REPEAT      537    564       4.
FT   REPEAT      565    592       5.
FT   REPEAT      593    620       6.
FT   REPEAT      621    648       7.
FT   REPEAT      649    676       8.
FT   REPEAT      677    704       9.
FT   REPEAT      705    732       10.
FT   REPEAT      733    760       11.
FT   REPEAT      761    788       12.
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1252   1252       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1293   1293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1342   1342       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1400   1400       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1802 AA;  188890 MW;  E344CA6469785A24 CRC64;

alignment_scores:
             Quality:  127.00          Length:    460
               Ratio:  0.547             Gaps:     22
 Percent Similarity:  50.435   Percent Identity: 22.609

alignment_block:
 US-09-303-518D-131/rev x HKR1_YEAST   ..

 Align seg 1/1  to: HKR1_YEAST  from: 1  to: 1802

 1337 TCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTCGTA 1288
       |||   :::::||||||::::::         |||:::   |||||||||
  390 SerAspThrValValSerThrSerSerThrAsnThrValProTyrSerSe 406

 1287 TTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAAT 1238
       :::       :::|||   :::|||   |||   ||||||||:::
  406 rVal...HisSerThrPheValHis...AlaThrSerSerSerThrTyrI 421

 1237 CCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCG 1188
        ::::::             :::   |||:::|||   :::::::::|||
  421 leSerSerSerLeuTyrSerSerProSerLeuSerAlaSerValSerSer 437

 1187 CGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGT 1138
       :::   :::   :::         |||      |||:::   |||   ||
  438 HisPheGlyValAlaProPheProSerAlaTyrIleSerPheSerSerVa 454

 1137 GCCGATC.........GGTACCATGGCGCGGTCGCCGCCGTTGACGGCTG 1097
       ||||:::          :::|||   :::   ||||||      ::::::|
  454 lProValAlaValSerSerThrTyrThrSerSerProSerAlaSerValV 471

 1096 TCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGC 1047
       |||||                              :::   |||   :::|||
  471 alValProSerAlaTyrAlaSerSerProSerValProValAlaValSer 487

 1046 GTGATGGAGTATTTGTCCGGCTGCGGCGCA................... 1017
                |||   |||:::   :::|||
  488 ...SerThrTyrThrSerSerProSerAlaProAlaAlaIleSerSerTh 503

 1016 .......ACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGG 974
              ::::::|||::::::   :::            |||:::   |||:
  503 rTyrThrSerSerProSerAlaProValAlaValSerSerThrTyrThrS 520

  973 AAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATC 924
       ::                                   ||||||:::||||||
  520 erSer.......................ProSerAlaProAlaAlaIle 528

  923 ...GCACCGTTCAATACCGAACCGGAAATC..............ACGCG 892
         :::   ::::::::::::|||:::                    :::
  529 SerSerThrTyrThrSerSerProSerAlaProValAlaValSerSerTh 545

  891 GTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTC.... 846
          ::::::|||   ::::::||||||:::      :::|||:::
  545 rTyrThrSerSerProSerAlaProAlaAlaIleSerSerThrTyrThrS 562

  845 ..........GCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACT 807
                 ||||||   :::|||            :::      |||
  562 erSerProSerAlaProValAlaVal.........SerSerThrTyrThr 575

  806 TGCAGGCCG......CCCAAGGCAACCACGCGCTCGGTATTCAGACGGCC 763
           |||      |||   |||   :::   :::            ||
  576 SerSerProSerAlaProAlaAlaIleSerSerThrTyrThrSerSerPr 592

  762 TGTTACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCC 713
       |         :::   |||:::|||::::::|||
  592 o.........SerAlaProValAlaValSerSer............... 600

  712 ACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCGTGCCA 663
       |||            :::|||::::::                   :::
  601 ..ThrTyrThrSerSerProSerAlaPro..............AlaAla 611
```

```
662 CTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATTGGC 613
    ::::::    :::        :::    |||::::::    |||::::::    ::
612 IleSerSerThrTyrThrSerSerProSerAlaProValAlaValSerSe 628

612 AGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTT 563
    :::::::::::::::      |||||||||||||||:::    :::
628 rThrTyrThrSerSerProSerAlaProAlaAlaIleSerSerThrTyrT 645

562 TACGTTCG.................GTCAGGCGGCTCAATACCAACAGG 531
          |||                 |||            |||:::
645 hrSerSerProSerValProValAlaValSerSerThrTyrThrSerSer 661

530 CCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGC 481
    |||           |||:::    |||::::::      |||      |||::
662 Pro.........SerAlaProAlaAlaIleSerSerThrTyrThrSerSe 675

480 AGCCAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGG 431
     :    |||       |||:::::::::    |||        ::::::    ||||
675 rProSerValProValAlaValSerSerThrTyrThrSerSerProSerA 692

430 CATCTACGGCAGGGATTTGCTGAACGGACGGGTGCGAAGCGCAGTCCAT 381
    ||    :::|||              :::::::::      |||:::
692 laProAlaAla.........IleSerSerThrTyrThrSerSerProSer 705

380 AAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGC 331
          |||       :::    :::        |||:::::::          ||
706 AlaPro......ValAlaValSerSerThrTyrThrSerSerProSerAl 720

330 CAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCAA 281
    |      |||:::                |||:::::::    :::|||    ||||||:
720 aProAlaAla.........IleSerSerThrTyrThrSerSerProSerA 734

280 CGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCG 231
    ::    ::::::::::::    :::               ||||||         |||
734 laProValAlaValSerSerThrTyrThrSerSerProSerAlaProAla 750

230 GCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTTT 181
    ||||||            :::::::::    :::::::::|||:::          ::
751 AlaIle.........SerSerThrTyrThrSerSerProSerAlaProVa 764

180 GTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTTGATTT 131
    ::::    :::|||||||:::               |||:::|||||||||    :::
764 lAlaValSerSerThrTyr.........ThrSerSerProSerAlaLeuV 778

130 TCATCGAGGGGCGCATGCCGACATATTCT..................TCG 99
      :::    :::            |||    |||                  |||
778 alValLeuSerSerThrSerThrSerSerProTyrAspIleValTyrSer 794

 98 CCAAGCAACGCGACTTCGGTAATGGCCGGG 69
    ||||||:::    :::::::::    :::|||
795 ProSerThrPheAlaAlaIleSerSerGly 804
```

seq_name: SwissProt_40:RNFC_RHOCA

```
seq_documentation_block:
ID   RNFC_RHOCA     STANDARD;      PRT;   519 AA.
AC   Q52716; O08057; Q52712;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Electron transport complex protein rnfC (Nitrogen fixation protein
DE   rnfC).
GN   RNFC.
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10S;
RX   MEDLINE=94088454; PubMed=8264535;
RA   Schmehl M., Jahn A., Meyer zu Vilsendorf A., Hennecke S., Masepohl B.,
RA   Schuppler M., Marxer M., Oelze J., Klipp W.;
RT   "Identification of a new class of nitrogen fixation genes in
RT   Rhodobacter capsulatus: a putative membrane complex involved in
RT   electron transport to nitrogenase.";
RL   Mol. Gen. Genet. 241:602-615(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.
RC   STRAIN=ATCC 33303 / B10;
RX   MEDLINE=98151232; PubMed=9492268;
RA   Jouanneau Y., Jeong H.-S., Hugo N., Meyer C., Willison J.C.;
RT   "Overexpression in Escherichia coli of the rnf genes from Rhodobacter
RT   capsulatus -- characterization of two membrane-bound iron-sulfur
RT   proteins.";
RL   Eur. J. Biochem. 251:54-64(1998).
RN   [3]
RP   SUBCELLULAR LOCATION.
RC   STRAIN=SB1003 / St Louis;
RX   MEDLINE=97299785; PubMed=9154934;
RA   Kumagai H., Fujiwara T., Matsubara H., Saeki K.;
RT   "Membrane localization, topology, and mutual stabilization of the
RT   rnfABC gene products in Rhodobacter capsulatus and implications for a
RT   new family of energy-coupling NADH oxidoreductases.";
RL   Biochemistry 36:5509-5521(1997).
CC   -!- FUNCTION: Required for nitrogen fixation. May be part of a
CC       membrane complex functioning as an intermediate in the electron
CC       transport to nitrogenase. Stabilizes rnfB.
CC   -!- COFACTOR: Binds 2 4FE-4S clusters (Potential).
CC   -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC       rnfD, rnfE and rnfG.
CC   -!- SUBCELLULAR LOCATION: Inner membrane-associated (Probable).
CC   -!- INDUCTION: Expression is reduced under iron-limiting conditions.
CC   -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC       RNFC SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X72888; CAA51399.1; -.
DR   EMBL; Y11913; CAA72670.1; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001949; Complex1_51K.
DR   Pfam; PF01512; Complex1_51K; 1.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Nitrogen fixation; Electron transport; Iron-sulfur; 4Fe-4S;
KW   Inner membrane.
FT   METAL       381    381       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       384    384       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       387    387       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       391    391       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       420    420       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       423    423       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       426    426       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       430    430       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   DOMAIN      498    503       POLY-ALA.
SQ   SEQUENCE   519 AA;  55587 MW;  E5451EB2A3FA6026 CRC64;


alignment_scores:
            Quality:  126.50          Length:    457
              Ratio:  0.586             Gaps:     18
 Percent Similarity:  47.265   Percent Identity: 18.381

alignment_block:
 US-09-303-518D-131 x RNFC_RHOCA   ..

 Align seg 1/1  to: RNFC_RHOCA  from: 1  to: 519

     70 CCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCG 119
```

```
        ||||||:::   ::::::  |||      ::::::::::::|||   ::
 42 ProAlaLeuIleArgLeuProLeu......GlnGlnHisIleGlyAlaGl 56

120 CCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGC 169
     :   :::      :::|||:::   :::   |||   |||||||||::::
 56 uAlaGluProIleValLysArgAspAspLeuValLeuLysGlyGlnLeuI 73

170 TGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCA 219
     ::   :::   :::       ::::::   :::   |||||||:::|||
 73 leAlaLysAlaArgGlyProLeuSerAlaAsnIleHisAlaProThrSer 89

220 GGCAAAATCGCCGCTATT.................CACCGTGGCGAAAA 251
    |||::::::   |||:::                  |||   :::
 90 GlyArgValIleAlaValGlyHisPheValAlaProHisAlaSerGlyLe 106

252 GCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAA.... 297
         |||       ::::::::|||      :::||||:::||||:::
106 uProVal...ProThrIleThrIleArgProAspGlyGluAspLysTrpG 122

298 .................ATCGAGTTCGAACGCTACGTACCTGAAGCGCTG 330
                      ::::::   |||:::   :::||||||   :::
122 lyProHisLeuProArgLeuArgProGluAsnAlaAlaProGluGluIle 138

331 GCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTT 380
    |||          ::::::   :::     :::::::::       |||
139 Ala.........AlaGlnValAlaAlaAlaGlyIleValGlyMetGlyGl 152

381 ATGGACT.........GCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTG 421
        |||          |||::::::   :::      :::|||
152 yAlaThrPheProSerAlaValLysLeuAsnLeuArgAlaLysTyrAspL 169

422 CCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAAT 471
      :::                      :::::::::::|||:::   :::   :::
169 euThr..............ThrLeuIleIleAsnGlyAlaGluCysGlu 180

472 CCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTT 521
    |||        :::          ::::::::|||   ||||||:::::
181 ProTyrLeuThrCysAspAspArgLeuMetArgGluArgAlaGluGluIl 197

522 CAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATG 571
    :        |||:::   :::::::::|||             |||    :::|
197 eAlaAspGlyIleGlyIleMetAlaArgAlaLeuGlyValLys...GlnV 213

572 TGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCC...... 615
    ||     |||        ::::::   |||:::          |||
213 alPheValAlaIleGluSerAsnLysProGlnAlaIleGluAlaMetThr 229

616 .......................AATATCGAAACACATGAATTTGGCGG 641
                            :::::::::   |||
230 ArgTyrAsnArgAlaLeuGlyTyrThrPheLysIleHisValValProTh 246

642 CCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAG 691
         :::|||           |||::::::   |||:::::::::    :
246 rGlnTyrPro.........MetGlySerGluLysHisLeuValLysMetI 260

692 TCGGCGCGAATAAAACC............................GTG 711
    ::    :::::::::|||                           |||
260 leThrGlyGlnGluThrProAlaArgAlaLeuThrAlaAspLeuGlyVal 276

712 TGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAAC 761
       :::   |||          :::   |||:::
277 ValValHisAsnIleAlaThrAlaHisAlaValHisLeuAlaValArgTy 293

762 AGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCA 811
     |||:::          |||:::|||:::::::::|||          ::::
293 rGlyGluProLeuIleAlaArgThrValThrValSerGlyHisGlyIleA 310

812 ACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTT 861
     ::::::|||    ||||||:::::::::|||:::   ||||||::::::
310 rgArgProAlaAsnLeuArgValLeuIleGlyThrProValSerGluIle 326

862 ACCGCC......GGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGG 905
       |||       |||   :::::::::       :::|||::::::   ||
327 IleAlaHisCysGlyGlyPheThrGluGluProAspArgLeuLeuLeuGl 343

906 TTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGAC 955
    |:::    :::     |||          |||
343 yGlyProMetMetGlyMetProIleGln...................... 352

956 GCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTG 1005
         :::   ::::::   |||::::::
353 .....AsnProArgValProValValLysGly.................. 361

1006 TTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCAC 1055

361 ................................................. 361

1056 TCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCA 1105
                       |||   ::::::   :::|||:::|||    :
362 ...............ThrAsnGlyIleLeuAlaLeuThrAlaAlaGluT 373

1106 ACGGCGGCGACCGCGCCATG......GTACCGATCGGCACTTATGAGCGC 1149
     ::               :::::::|||          :::    |||         :::
373 hrProGluAlaLysThrMetProCysIleArgCysGlyArgCysValGln 389

1150 GTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGT 1199
        |||:::   :::   |||      |||          :::    ::
390 GlyCysProValGlyLeuThrProPheGluLeuAsnAlaArgIleHisAl 406

1200 CGGCGATACCGACAGCGCGCAGGCTTTGGGT.........TGCTTGGAAT 1240
     :||||||    ::::::|||      :::|||          ||||||
406 aGlyAspLeuGluGlyAlaAlaLysValGlyLeuMetAspCysLeuAla. 422

1241 TGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATAC 1290
                       :::   ||||||:::   ||||||::::::
423 ..............CysGlyCysCysSerTyrAsnCysProAlaAsnLeu 434

1291 GAATACGGCCCGCTGTTGCGC 1311
             ||||||::::::
435 .........ProLeuValGln 438
```

seq_name: SwissProt_40:RNFC_ECO57

```
seq_documentation_block:
ID   RNFC_ECO57     STANDARD;      PRT;   740 AA.
AC   P58324;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Electron transport complex protein rnfC.
GN   RNFC OR Z2636 OR ECS2338.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
```

```
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: May be part of a membrane complex involved in electron
CC       transport (By similarity).
CC   -!- COFACTOR: Binds 2 4FE-4S clusters (Potential).
CC   -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC       rnfD, rnfE and rnfG (By similarity).
CC   -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC       RNFC SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE005386; AAG56618.1; -.
DR   EMBL; AP002558; BAB35761.1; -.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane.
FT   METAL       377    377       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       380    380       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       383    383       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       387    387       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       416    416       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       419    419       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       422    422       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       426    426       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   CONFLICT    651    651       R -> RKAAVEAAIARAKARKLEQQQANAEPEEPVDPR
FT                                (IN REF. 2).
SQ   SEQUENCE   740 AA;  80297 MW;  8EE8A9E22DEC56F8 CRC64;

alignment_scores:
          Quality:  121.00             Length:    319
            Ratio:  0.742                Gaps:     14
 Percent Similarity:  51.097   Percent Identity:  22.571

alignment_block:
 US-09-303-518D-131 x RNFC_ECO57   ..

 Align seg 1/1   to: RNFC_ECO57   from: 1   to: 740

     100 GAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGC 149
          :::::::::::|||   :::   ::::::   ::::::   |||:::
      48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspLy 64

     150 CGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAG 199
            |||    :::||||||   |||   :::   :::   :::          |
      64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuProV 81

     200 TATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATT............ 237
           ||       ||||||:::||||||   ::::::||||||
      81 alHis...AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96

     238 ......CACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
                 |||    :::              |||   |||||||:::|||    ::
      97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112

     282 TGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCT...GAAGCGC 328
          ::::|||:::|||                  ::::::|||    ::::::
     112 aAspGlyGluAsp..............CysTrpIleProArgAspGlyT 124

     329 TGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAA...... 372
            |||         |||::::::   |||::::::||||||:::
     124 rpAlaAspTyrArgSerArgArg...ArgGluGluLeuIleGluArgIle 139

     373 ..............................TCAGGCTTATGGACTGCGCT 392
                                       :::|||:::   |||::::::
     140 HisGlnPheGlyValAlaGlyLeuGlyGlyAlaGlyPheProThrGlyVa 156

     393 TCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCG 442
          :                 |||:::   :::            :::     :
     156 l..................LysLeuGlnGlyGlyGlyAspLysIleGluT 167

     443 CCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACG 492
          :::::::::::||||||   :::   :::|||         :::
     167 hrLeuIleIleAsnAlaAlaGluCysGluProTyrIleThrAlaAspAsp 183

     493 GTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATT 542
            ::::::::::::::|||   :::         :::|||:::   :::||
     184 ArgLeuMetGlnAspCysAlaAlaGlnValValGluGlyIleArgIleLe 200

     543 GAGCCGCCTGACCGAA.................................. 558
          |:::::::::   :::
     200 uAlaHisIleLeuGlnProArgGluIleLeuIleGlyIleGluAspAsnL 217

     559 .....CGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCT 603
               :::   |||   :::   :::|||:::   ||||||
     217 ysProGlnAlaIleSerMetLeuArgAlaValLeuAlaAsp......... 230

     604 GAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGC 653
               :::   :::|||:::   :::   :::            :::|||::
     231 ......SerHisAspIleSerMetArgValIleProThrLysTyrProSe 245

     654 CGGC.......................TTGAGTGGCACGCACATTCATT 679
          :|||                         |||:::|||   ::::::
     245 rGlyGlyAlaLysGlnLeuThrTyrIleLeuThrGlyLysGlnVal.... 260

     680 TCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAA 729
                 |||   |||::::::::::::         :::         |||
     261 ........ProHisGlyGlyArgSerSerAspIleGlyValLeuMetGln 274

     730 GACGTG.........ATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCT 770
          :::|||            |||:::   |||         :::   |||:::
     275 AsnValGlyThrAlaTyrAlaValLysArgAlaValIleAspGlyGluPr 291

     771 GAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGC 820
              |||||||||||||||:::|||   |||         :::   :::|||
     291 oIleThrGluArgValValThrLeuThrGlyGluAlaIleAlaArgProG 308

     821 GCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTT......ACC 864
              :::   :::   ||||||:::   |||      :::|||
     308 lyAsnValTrpAlaArgLeuGlyThrProValArgHisLeuLeuAsnAsp 324

     865 GCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATT 914
          ||||||           :::||||||:::   ||||||   |||:::   ||
     325 AlaGlyPheCysProSerAlaAspGlnMetValIleMetGlyGlyProLe 341

     915 GAACGGT 921
          |   |||
     341 uMetGly 343

seq_name: SwissProt_40:RNFC_ECOLI

seq_documentation_block:
ID   RNFC_ECOLI     STANDARD;      PRT;   740 AA.
AC   P77611;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Electron transport complex protein rnfC.
GN   RNFC OR B1629.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
```

```
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
CC   -!- FUNCTION: May be part of a membrane complex involved in electron
CC       transport (By similarity).
CC   -!- COFACTOR: Binds 2 4FE-4S clusters (Potential).
CC   -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC       rnfD, rnfE and rnfG (By similarity).
CC   -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC       RNFC SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE000258; AAC74701.1; -.
DR   EMBL; D90806; BAA15384.1; -.
DR   EMBL; D90807; BAA15391.1; -.
DR   EMBL; D90808; BAA15414.1; -.
DR   HSSP; P00198; 1FCA.
DR   EcoGene; EG13935; rnfC.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001949; Complex1_51K.
DR   Pfam; PF01512; Complex1_51K; 1.
DR   Pfam; PF00037; fer4; 2.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW   Complete proteome.
FT   METAL       377    377       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       380    380       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       383    383       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       387    387       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       416    416       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       419    419       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       422    422       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       426    426       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   740 AA;  80171 MW;  381506475CD6E01F CRC64;

alignment_scores:
          Quality:  117.00         Length:     313
            Ratio:  0.731            Gaps:      13
 Percent Similarity:  51.118   Percent Identity:  22.684

alignment_block:
 US-09-303-518D-131 x RNFC_ECOLI  ..

 Align seg 1/1  to: RNFC_ECOLI  from: 1  to: 740

    100 GAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGC 149
        :::::::::::|||   :::   ::::::   ::::::   |||:::
     48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspLy 64

    150 CGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAG 199
         |||   :::|||||||   |||   :::   :::   :::      |
     64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuProV 81

    200 TATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATT............ 237
         ||         ||||||:::||||||   ::::::||||||
     81 alHis...AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96

    238 ......CACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
              |||   :::            |||   ||||||:::|||   ::
     97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112

    282 TGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAA....... 324
        :::|||:::|||                 ::::::|||:::
    112 aAspGlyGluAsp..............CysTrpIleProArgAspGlyT 124

    325 .....GCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATT 369
                          |||   |||::::::   :::::::::
    124 rpAlaAspTyrArgThrArgSerArgGluGluLeuIleGluArgIleHis 140

    370 CAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGC........ 411
        |||   |||:::   ::::::|||               |||
    141 GlnPheGlyVal...AlaGlyLeuGlyGlyAlaGlyPheProThrGlyVa 156

    412 .AAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGA 460
         |||:::   :::            :::      :::::::::::||||||
    156 lLysLeuGlnGlyGlyGlyAspLysIleGluThrLeuIleIleAsnAlaA 173

    461 TGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCC 510
         :::   :::|||         :::            ::::::::::::::
    173 laGluCysGluProTyrIleThrAlaAspAspArgLeuMetGlnAspCys 189

    511 GCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAA.. 558
        |||   :::      :::|||:::   :::|||:::::::::   :::
    190 AlaAlaGlnValValGluGlyIleArgIleLeuAlaHisIleLeuGlnPr 206

    559 ....................................CGTAAAATCCATG 571
                                            :::   |||   :
    206 oArgGluIleLeuIleGlyIleGluAspAsnLysProGlnAlaIleSerM 223

    572 TGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATC 621
        ::   :::|||:::   ||||||              :::   :::|||
    223 etLeuArgAlaValLeuAlaAsp..............SerAsnAspIle 234

    622 GAAACACATGAATTTGGCGGCCCGCATCCTGCCGGC.............. 657
        :::   :::   :::      :::|||:::|||
    235 SerLeuArgValIleProThrLysTyrProSerGlyGlyAlaLysGlnLe 251

    658 .........TTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCG 697
                 |||:::|||   ::::::            |||   |||:
    251 uThrTyrIleLeuThrGlyLysGlnVal...........ProHisGlyG 264

    698 CGAATAAAACCGTGTGGACCATCAATTATCAAGACGTG.........ATT 738
        ::::::::::        :::      |||:::|||
    264 lyArgSerSerAspIleGlyValLeuMetGlnAsnValGlyThrAlaTyr 280

    739 GCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGT 788
        |||:::   |||         :::   |||:::      |||||||||||||
    281 AlaValLysArgAlaValIleAspGlyGluProIleThrGluArgValVa 297

    789 TGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTT 838
        |:::|||   |||         :::   :::|||      :::   :::   |
    297 lThrLeuThrGlyGluAlaIleAlaArgProGlyAsnValTrpAlaArgL 314

    839 TGGGTGCGAAGGTGTCTCAACTT......ACCGCCGGCGAATTGGTTGAC 882
        ||||||:::   |||   :::|||          ||||||         :::
    314 euGlyThrProValArgHisLeuLeuAsnAspAlaGlyPheCysProSer 330

    883 GCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGT 921
        |||||||:::   ||||||   |||:::   |||   |||
    331 AlaAspGlnMetValIleMetGlyGlyProLeuMetGly 343

seq_name: SwissProt_40:ODP2_MYCCA
```

```
seq_documentation_block:
ID   ODP2_MYCCA     STANDARD;      PRT;   438 AA.
AC   Q49110;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase
DE   complex (EC 2.3.1.12) (E2).
GN   PDHC OR ODP2.
OS   Mycoplasma capricolum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Entomoplasmataceae.
OX   NCBI_TaxID=2095;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97001869; PubMed=8844861;
RA   Zhu P.P., Peterkofsky A.;
RT   "Sequence and organization of genes encoding enzymes involved in
RT   pyruvate metabolism in Mycoplasma capricolum.";
RL   Protein Sci. 5:1719-1736(1996).
CC   -!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC       CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC       COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC       DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC       (E3) (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide = CoA + S-
CC       acetyldihydrolipoamide.
CC   -!- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
CC       COFACTOR (BY SIMILARITY).
CC   -!- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
CC       SYMMETRY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U62057; AAC44344.1; -.
DR   HSSP; P11961; 1B5S.
DR   InterPro; IPR001078; 2Oxoacid_dh.
DR   InterPro; IPR000089; Biotin_lipoyl.
DR   InterPro; IPR003016; Lipoyl.
DR   InterPro; IPR004167; e3_binding.
DR   Pfam; PF00198; 2-oxoacid_dh; 1.
DR   Pfam; PF00364; biotin_lipoyl; 1.
DR   Pfam; PF02817; e3_binding; 1.
DR   ProDom; PD001115; 2Oxoacid_dh; 1.
DR   PROSITE; PS00189; LIPOYL; 1.
KW   Glycolysis; Transferase; Acyltransferase; Lipoyl.
FT   BINDING      42     42       LIPOYL (BY SIMILARITY).
FT   ACT_SITE    411    411       POTENTIAL.
SQ   SEQUENCE   438 AA;  46927 MW;  4BF83B697480B4AB CRC64;

alignment_scores:
           Quality:  116.50            Length:    457
             Ratio:  0.574               Gaps:     21
 Percent Similarity:  44.420   Percent Identity: 20.131

alignment_block:
 US-09-303-518D-131 x ODP2_MYCCA   ..

 Align seg 1/1  to: ODP2_MYCCA  from: 1  to: 438

      67 GGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCAT 116
          |||    ::::::||||||                          |||
      10 GlyGluGlyLeuThrGlu..........................GlyTh 17

     117 GCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAG 166
                :::::::    :::|||   |||::::::||||||:::||||||
      17 rValAlaGluValLeuValLysValGlyAspValValLysGluGlyGlnS 34

     167 TGCTGTTT.........GAAGACAAAAAGAATCCGGGCGTAGTATTTACT 207
           |||:::                ||||||  |||       :::
      34 erLeuTyrPheValGluThrAspLysValAsnSerGluIle......Pro 48

     208 GCGCCGGCTTCAGGCAAAATCGCCGCTATTCAC......CGTGGCGAAAA 251
         ||||||:::::||||||||||||:::|||:::            |||:::::
      49 AlaProValAlaGlyLysIleAlaValIleAsnIleLysAlaGlyGlnGl 65

     252 GCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAA...GGCAACGACGAAA 298
         :        :::    :::|||||||:::    :::|||    |||:::|||
      65 uIleLysValGlyAspValValMetGluIleGluAspGlySerAspThrS 82

     299 TCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAA 348
                ::::::               |||:::   ::::::   :::   |||
      82 erAlaThrSerGluProLysAlaGluThrLysSerGluAlaLysValGlu 98

     349 AAAGTGCGCCGCAACCTG....................ATTCAATCAGG 377
           |||:::::|||                          :::::::::
      99 ValValGluGluAsnAlaSerValValGlyAlaThrProValSerAsnAs 115

     378 CTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAG 427
          :::    :::        :::|||           :::|||        ::::::
     115 pValIleValArgLysGlnThrThrThrValAsnLysSerSerThrIleL 132

     428 ATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTG 477
            |||    |||:::|||                                :::    :::
     132 ysAlaThrProLeuAla......................ArgLysVal 140

     478 GCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACG 527
         |||||||||    ::::::    ::::::
     141 AlaAlaAspLeuAsnIleAspLeuSer...................... 149

     528 CGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTA 577
                  ||||||                  :::::::::|||     |||
     150 .......LeuValThrProThrGlyProAsnGlnArgIleLeuVal.... 162

     578 AAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAAT......ATC 621
                    ||||||:::    :::::::::|||::::::          :::
     163 ...........AlaAspIleLysAsnHisGlnAlaSerSerThrGlnLeu 175

     622 GAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCA 671
            ::::::   ::::::   |||    |||:::               ::::::||
     176 AlaSerGlnProIleSerGlnProAlaProThrProSerProSerAlaHi 192

     672 C.................ATTCATTTCATCGAGCCAGTCGGC....... 696
         |                  |||      :::||||||     :::
     192 sGlnThrIleAlaProThrIleLysValValGluProSerAlaProLeuS 209

     697 .................................GCGAATAAAACCGTG 711
                                            |||:::       :::
     209 erTrpAspGluValProMetAsnGlyValArgLysAlaThrValLysAla 225

     712 TGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAAC 761
            |||    ::::::    ::::::    |||:::                  ||
     226 MetThrLysSerHisThrGluIleAlaAlaPhe..............Th 237

     762 AGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCA 811
         ||||      ||||||:::                                 ::::
     237 rGlyMetLysAsnThrAsp..........................IleT 245

     812 ACAAACCGCGCCTCTTGCGTACCGTTTTG..................... 840
         :::::    :::    :::||||||    |||
     245 hrGluThrHisLysMetArgThrGluLeuLysAspHisAlaAlaAlaSer 261

     841 GGTGCGAAGGTGTCTCAACTT.................ACCGCCGGCGA 872
         |||    |||::::::    |||                  :::|||    ::
     262 GlyIleLysLeuThrTyrLeuAlaPheIleIleLysAlaValAlaLysSe 278

     873 ATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGTG 922
```

```
        :|||   |||     |||      :::  |||:::     |||:::
 278 rLeuArgAspMetProAsnIleAsnValArgGlyAspPheAlaAsnAsnL 295

 923 CGATT............................................. 927
       |||
 295 ysIleGlnPheMetHisAsnIleAsnIleGlyIleAlaValAspThrPro 311

 928 .....................GCACAAGGCGCGCATGATTATTTGGGACG 956
                          :::||||||   |||:::|||:::
 312 AsnGlyLeuMetValProValIleLysGlyAla...AspHisLeuSerVa 327

 957 CTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGT 1006
      ::::::      |||::::::   |||
 327 lPheGluIleAlaIleLysIleSerGlu...................... 336

1007 TCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACC... 1053
             :::|||   :::            ::::::||||||:::
 337 ........LeuAlaAsnLysAlaLysAspGlyLysLeuThrArgAlaGlu 350

1054 ..................ACTCTCGGCCATTTCCTAAAAAACAAACTCTT 1085
                       |||:::::::::|||   :::      |||
 351 MetThrGluAlaThrPheThrValSerAsnPheGlySerValGlyLeuAs 367

1086 CAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCG 1135
              |||      :::|||:::   :::   |||::::::    :::|
 367 pTyrAlaThrProIleIleAsnSerProGluSerAlaIleLeuGlyValG 384

1136 GCACTTATGAGCGC.................................... 1149
      |||||   ::::::
 384 lyThrMetSerGlnThrProLeuTyrIleAsnGlyGluLeuGlnLysArg 400

1150 ...GTAATGCCGTTGGACATC 1167
        :::|||||||||::::::
 401 PheIleMetProLeuSerMet 407
```

seq_name: SwissProt_40:FAS1_SCHPO

```
seq_documentation_block:
ID   FAS1_SCHPO     STANDARD;      PRT;  2073 AA.
AC   Q9UUG0; P78799;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fatty acid synthase subunit beta (EC 2.3.1.86) [Includes: 3-
DE   hydroxypalmitoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.61);
DE   Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9); [Acyl-
DE   carrier-protein] acetyltransferase (EC 2.3.1.38); [Acyl-carrier-
DE   protein] malonyltransferase (EC 2.3.1.39); S-acyl fatty acid synthase
DE   thioesterase (EC 3.1.2.14)].
GN   FAS1 OR SPAC926.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., AND SUBUNIT.
RX   MEDLINE=98359895; PubMed=9693066;
RA   Niwa H., Katayama E., Yanagida M., Morikawa K.;
RT   "Cloning of the fatty acid synthetase beta subunit from fission yeast,
RT   coexpression with the alpha subunit, and purification of the intact
RT   multifunctional enzyme complex.";
RL   Protein Expr. Purif. 13:403-413(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Wood V., Rajandream M.A., Barrell B.G., Ramsperger U., Pohl T.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1725-2073 FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE=98162722; PubMed=9501991;
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
RL   DNA Res. 4:363-369(1997).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THE BETA SUBUNIT CONTAINS DOMAINS FOR: [ACYL-CARRIER PROTEIN]
CC       ACETYLTRANSFERASE AND MALONYLTRANSFERASE, S-ACYL FATTY ACID
CC       SYNTHASE THIOESTERASE, ENOYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND
CC       3-HYDROXYPALMITOYL-[ACYL-CARRIER PROTEIN] DEHYDRATASE.
CC   -!- CATALYTIC ACTIVITY: ACETYL-COA + N MALONYL-COA + 2N NADPH =
CC       LONG-CHAIN FATTY ACID + (N+1) COA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NAD(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: FATTY ACID SYNTHETASE ARE [ALPHA(6)BETA(6)] HEXAMERS OF
CC       TWO MULTIFUNCTIONAL SUBUNITS (ALPHA & BETA).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB010274; BAA36384.1; -.
DR   EMBL; AL110469; CAB54157.1; -.
DR   EMBL; D89148; BAA13810.1; -.
DR   InterPro; IPR001227; Acyltransf_domain.
DR   InterPro; IPR003965; Fatty_acid_synth.
DR   InterPro; IPR002539; MaoC_dehydratas.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF01575; MaoC_dehydratas; 1.
DR   PRINTS; PR01483; FASYNTHASE.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; Hydrolase; Lyase; NAD; NADP.
FT   DOMAIN        1    459       ACETYL TRANSFERASE.
FT   DOMAIN      470    858       ENOYL REDUCTASE.
FT   DOMAIN     1155   1644       DEHYDRATASE.
FT   DOMAIN     1645   2073       MALONYL/PALMITOYL TRANSFERASE.
FT   ACT_SITE    270    270       ACETYL TRANSFERASE (BY SIMILARITY).
FT   ACT_SITE   1361   1361       DEHYDRATASE (POTENTIAL).
FT   ACT_SITE   1828   1828       MALONYLTRANSFERASE (BY SIMILARITY).
FT   CONFLICT    222    222       P -> R (IN REF. 1).
SQ   SEQUENCE   2073 AA;  230559 MW;  D928270838E7C7C4 CRC64;

alignment_scores:
            Quality:  114.50         Length:    461
              Ratio:  0.556            Gaps:     25
 Percent Similarity:  44.685   Percent Identity: 21.475

alignment_block:
 US-09-303-518D-131 x FAS1_SCHPO   ..

 Align seg 1/1  to: FAS1_SCHPO  from: 1  to: 2073

 217 TCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAGCGCGTACTTCAGTC 266
      :::||||||:::      ::::::|||   |||:::|||      |||::
 948 ThrGlyLysLeu......MetArgArgIleGluGluArgPheThrGlnAs 962

 267 AGTC......GTGATTGCCGTTGAAGGC...........AACGACGAAA 298
      :|||       :::       :::|||:::           ||||||
 962 pValGlyLysThrThrLeuIleGluAsnPheGluAspLeuAsnAspProT 979

 299 TCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAA 348
```

```
                                  :::|||      :::   |||::::::
 979 yrProValAlaAlaArgPheLeuAspAlaTyrProGluAlaSerThrGln 995

 349 AAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCAC 398
        :::     |||     ::::::   ::::::   :::|||   ::
 996 AspLeu......AsnThrGlnAspAlaGlnPhePheTyrSerLeuCysSe 1010

 399 CCGTCCGTTCAGCAAA...........ATCCCTGCCGTAGAT....... 429
     ::::||||||:::|||                |||||||||:::|||
1010 rAsnProPheGlnLysProValProPheIleProAlaIleAspAspThrP 1027

 430 ................................GCCGAGCCGTTCGCC 444
                                     :::|||   :::|||
1027 heGluPheTyrPheLysLysAspSerLeuTrpGlnSerGluAspLeuAla 1043

 445 ATCTTCGTCAAT..................GCGATGGACACCAATCCGCT 476
                |||:::              |||:::      :::|||::
1044 AlaValValGlyGluAspValGlyArgValAlaIleLeuGlnGlyProMe 1060

 477 GGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAAC 526
     :||||||      ::::::   :::::::|||   |||::::::
1060 tAlaAlaLysHisSerThrLysValAsnGluProAlaLysGlu....... 1074

 527 GCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATC...CATGTG 573
                 :::|||:::    :::::::|||     |||    |||
1075 ..........LeuLeuAspGlyIleAsnGluThrHisIleGlnHisPhe 1087

 574 TGTAAAGCA......GCAGGCGCAGAC......GTGCCGTCTGAAAATGC 611
        |||           ||||||    :::        :::|||
1088 IleLysLysPheTyrAlaGlyAspGluLysLysIleProIle........ 1101

 612 TGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGA 661
                 ::::::   |||||||||     |||:::::::::|
1102 ............ValGluTyrPheGlyGlyValProProValAsnValS 1114

 662 GT....................................GGCACGCAC 672
     ||                                    |||:::
1114 erHisLysSerLeuGluSerValSerValThrGluGluAlaGlySerLys 1130

 673 ATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTG........... 711
     ::::::   :::      :::|||:::|||:::::::::
1131 ValTyrLysLeuProGluIleGlySerAsnSerAlaLeuProSerLysLy 1147

 711 ................................................ 711

1147 sLeuTrpPheGluLeuLeuAlaGlyProGluTyrThrTrpPheArgAlaI 1164

 712 ............................TGGACCATCAATTATCAAGAC 732
                                 |||   ::::::::::::
1164 lePheThrThrGlnArgValAlaLysGlyTrpLysLeuGluHisAsnPro 1180

 733 GTGATTGCTATC..................GGACGTTTGTTCGTAACAGG 764
     |||       |||                   |||        |||   ||
1181 ValArgArgIlePheAlaProArgTyrGlyGlnArgAlaValValLysGl 1197

 765 CCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACA 814
     |:::   |||    :::    ||||||    |||        |||   :::
1197 yLysAspAsn...AspThrValValGluLeuTyrGluThrGlnSerGly. 1212

 815 AACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACC 864
                  :::   ||||||:::|||:::::::|||
1213 .............AsnTyrValLeuAlaAlaArgLeuSer......... 1221

 865 GCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATT 914
                   |||:::::::::   |||:::|||
1222 ...........TyrAspGlyGluThrIleValValSerMetPheGluAs 1234

 915 GAACGGTGCGATTGCACAAGGCGCGCAT...GATTATTTGGGACGCTAC. 960
       ::::::|||:::    :::     :::|||    |||:::|||    :::|||
1234 nArgAsnAlaLeuLysLysGluValHisLeuAspPheLeuPheLysTyrG 1251

 961 .............CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGC 996
                  ::::::    :::|||    |||    :::||||||:::
```

```
1251 luProSerAlaGlyTyrSerProValSerGluIleLeuAspGlyArgAsn 1267

 997 AAAGAGCTG......TTCGGCTGGGTT..............GCGCCGCA 1025
       ::::::        |||   |||:::                  |||
1268 AspArgIleLysHisPheTyrTrpAlaLeuTrpPheGlyGluGluProTy 1284

1026 GCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAA. 1074
       |||:::::::   |||||||||    |||       |||      :::
1284 rProGluAsnAlaSerIleThrAspThrPheThrGlyProGluValThrV 1301

1075 ........AACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGAC 1116
                 |||    :::       |||     :::|||::::::   :::
1301 alThrGlyAsnMetIleGluAspPheCysArgThrValGlyAsnHisAsn 1317

1117 CGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGAC.. 1164
       :::|||   :::          :::      :::|||:::    |||:::|||
1318 GluAlaTyrThrLysArgAlaIleArgLysArgMetAlaProMetAspPh 1334

1165 ...................................ATCCTGCCTACCT 1177
                                        |||:::|||
1334 eAlaIleValValGlyTrpGlnAlaIleThrLysAlaIlePheProLysA 1351

1178 TGCTTTTGCGCGATTTA......ATCGTCGGCGATACCGACAGCGCGCAG 1221
        :::       ||||||       :::|||       ::::::|||   :::
1351 laIleAspGlyAspLeuLeuArgLeuValHisLeuSerAsnSerPheArg 1367

1222 GCTTTGGGTTGCTTGGAATTGGACGAAGAAGAC 1254
        :::|||      :::|||    |||    |||
1368 MetValGlySerHisSerLeuMetGluGlyAsp 1378
```

seq_name: SwissProt_40:PPSA_PYRAB

```
seq_documentation_block:
ID   PPSA_PYRAB     STANDARD;      PRT;   819 AA.
AC   Q9V2H7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate,water
DE   dikinase) (PEP synthase).
GN   PPSA OR PAB0057.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORSAY;
RA   Heilig R.;
RT   "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT   structure and evolution.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC       phosphoenolpyruvate + phosphate.
CC   -!- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC       LACTATE ARE USED AS A CARBON SOURCE.
CC   -!- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ248283; CAB49021.1; -.
DR   HSSP; P22983; 1DIK.
DR   InterPro; IPR000121; PEP_utilizers.
DR   InterPro; IPR002192; PPDK_N_term.
DR   Pfam; PF00391; PEP-utilizers; 1.
DR   Pfam; PF02896; PEP-utilizers_C; 1.
DR   Pfam; PF01326; PPDK_N_term; 1.
DR   ProDom; PD000940; PEP_utilizers; 1.
```

```
DR   PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR   PROSITE; PS00742; PEP_ENZYMES_2; 1.
KW   Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.
FT   MOD_RES      441    441       PHOSPHORYLATION (BY SIMILARITY).
FT   DOMAIN       809    816       POLY-GLU.
SQ   SEQUENCE   819 AA;  90652 MW;  E09486ABB527C6EA CRC64;

alignment_scores:
            Quality:  113.00            Length:     475
              Ratio:  0.467               Gaps:      20
 Percent Similarity:  50.947   Percent Identity:  20.421

alignment_block:
 US-09-303-518D-131 x PPSA_PYRAB  ..

 Align seg 1/1  to: PPSA_PYRAB  from: 1  to: 819

      31 CCCATCGCGGGCAGACCGGAGCAAGTCATTTATGAC.............. 66
          |||::::::::::::   ::::::::::::   :::
     235 ProValThrAsnAsnArgAsnGluIleMetIleAsnAlaSerTrpGlyLe 251

      67 .GGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCA 115
           |||   |||:::::::::   |||:::      :::|||||||:::
     251 uGlyGluAlaValValSerGlyAlaValThrProAspGluTyrIleValG 268

     116 TGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAA 165
            :::    :::    ||||||||||||    :::::::::    |||    :::
     268 luLysGlyThrTrpLysIleLysGlu...LysValIleAlaLysLysGlu 283

     166 GTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGC 215
          |||:::                        |||:::      |||
     284 ValMet.........................ValIleArgAsnProGl 291

     216 TTCAGGCAAA...............ATCGCCGCTATTCACCGTGGC.... 246
          :::||||||                 :::|||    :::    |||
     291 uThrGlyLysGlyThrValThrValLysValAlaGluTyrLeuGlyProG 308

     247 ........GAAAAGCGCGTACTT.........CAGTCAGTCGTGATTGCC 279
                 ||||||:::||||||         |||    :::    :::|||
     308 luTrpValGluLysGlnValLeuThrAspGluGlnIleIleGluValAla 324

     280 GTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCT 329
                |||:::            |||          :::       :::
     325 LysMetGlyGlnLysIleGluGluHisTyrGlyTrpProGlnAspIleGl 341

     330 GGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCT 379
                    :::::::::       :::       ::::::||||||
     341 uTrpAlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerArgP 358

     380 TATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGAT 429
               |||:::|||:::                :::::::::        |||
     358 roValThrThrLeuLysGluThrThrThrGluGluValGluGluVal... 373

     430 GCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG...CT 476
           |||          ::::::::::::::::::     ::::::|||     ::
     374 ...GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyIl 389

     477 GGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAAC 526
          :::::|||         :::|||:::|||      :::|||:::|||
     389 eGlyAlaGlyArgValValValIlePheAspAlaSerGlu.......... 402

     527 GCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGT 576
                         ::::::::::::    |||       :::    |||
     403 ..............IleAspLysValLysGluGlyAspValLeuValThr 414

     577 AAAGCAGCAGGCGCAGAC...GTGCCGTCT...GAAAATGCTGCCAATAT 620
               ::::::    |||    |||||||:::    ::::::||||||    ||
     415 ThrMetThrAsnProAspMetValProAlaMetLysArgAlaAlaAlaIl 431

     621 CGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGC 670
          |    |||    |||    |||           |||::::::    :::    :
     431 eIleThrAspGluGlyGlyArgThrSerHisAlaAlaIleValSerArgG 448

     671 ACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATC 720
          :::::    :::          ||||||::::::::::::::::       :::
     448 luLeuGlyIleProAlaValValGlyThrLysGluAlaThrLysLysLeu 464

     721 AATTATCAAGACGTGATTGCTATCGGA....................... 747
          :::        |||    :::::::::
     465 LysThrGlyAspTyrValThrValAspGlyThrArgGlyLeuValTyrLy 481

     748 .........CGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGG 787
                      :::    :::|||       :::    :::    |||    :::
     481 sGlyIleValLysSerLeuValGluLysLysLysLysGluGluAlaAlaA 498

     788 TTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTT 837
          :::|||    |||:::       :::       |||    :::       |||:::
     498 laAlaProGlyAlaAlaValAlaAlaAlaProLeuValThrGlyThrLeu 514

     838 TTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGA 887
          :::    ::::::|||||||:::          ::::::|||
     515 ValLysValAsnValSerMetProGluValAlaGlu.............. 526

     888 CAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCG 937
              |||:::    :::    ::::::    :::    ::::::    ::::::
     527 ....ArgAlaAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAlaG 542

     938 CGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAA 987
            |||       |||:::       ::::::    ::::::    |||:::|||
     542 luHisMetIleLeuSerIleGlyGlnHisProValLysPheIleLysGlu 558

     988 GGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATA 1037
          |||:::::::::||||||          :::|||          :::|||
     559 GlyLysGluGluGluLeuValGluLysLeuAlaGluGlyIleGluLysVa 575

    1038 CTCCATCACGCGCACCACTCTCGGCCATTT.................... 1067
           :::     |||     :::     ||||||
     575 lAlaAla..AlaPheTyrProArgProValTrpTyrArgThrLeuAspAl 591

    1068 .CCTAAAAAACAAACTCTTCAAGTT.........CACGACAGCCGTCAAC 1107
           |||    ::::::       ::::::              :::|||:::    ::::
     591 aProThrAsnGluPheArgGluMetProGlyGlyGluAspGluProGluG 608

    1108 GGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCC 1157
          :::|||:::|||       |||                |||::::::
     608 luArgAsnProMetLeuGlyTrpArgGly.........IleArgArgGly 621

    1158 GTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATA 1207
          :::    :::|||       |||       |||:::|||:::       :::
     622 LeuAspGlnProGluLeuLeuArg...AlaGluPheLysAlaIleLysLy 637

    1208 CCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTC 1257
                :::    |||::::::    :::|||
     637 sValValGluLysGlyTyrAsnAsnIleGly................... 647

    1258 GCTTTGTGCAGCTTCGTCTG.........CCCGGGCAAATACGAATACGG 1298
               |||:::|||   ·|||        |||    |||||||||
     648 .....ValMetLeuProLeuValSerHisProGluGlnIleArgLysAla 662

    1299 CCCGCTGTTGCGCAAAGTGCTGGA 1322
                :::|||:::|||:::|||
     663 LysGluIleAlaArgSerValGly 670

seq_name: SwissProt_40:CHLN_SYNP7

seq_documentation_block:
ID   CHLN_SYNP7     STANDARD;      PRT;   466 AA.
AC   P54208;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Light-independent protochlorophyllide reductase subunit N
DE   (EC 1.18.-.-) (LI-POR subunit N) (DPOR subunit N).
GN   CHLN OR FRXC.
OS   Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
```

```
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=1140;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95388763; PubMed=7659748;
RA   Ronen-Tarazi M., Lieman-Hurwitz J., Gabay C., Orus M.I., Kaplan A.;
RT   "The genomic region of rbcLS in Synechococcus sp. PCC 7942 contains
RT   genes involved in the ability to grow under low CO2 concentration and
RT   in chlorophyll biosynthesis.";
RL   Plant Physiol. 108:1461-1469(1995).
CC   -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
CC       protochlorophyllide (Pchlide) to form chlorophyllide a (Chlide)
CC       (By similarity). This reaction is light-independent.
CC   -!- PATHWAY: Light-independent chlorophyll biosynthesis.
CC   -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC       of three subunits; chlL, chlN and chlB. Could form a
CC       heterotetramer of two chlB and two chlN subunits (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE BCHN / CHLN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X67694; CAA47924.1; -.
KW   Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase.
SQ   SEQUENCE   466 AA;  51540 MW;  9B1A51305045574C CRC64;

alignment_scores:
           Quality:  112.50             Length:    471
             Ratio:  0.577                Gaps:     25
 Percent Similarity:  41.401   Percent Identity:  20.807

alignment_block:
 US-09-303-518D-131 x CHLN_SYNP7   ..

 Align seg 1/1   to: CHLN_SYNP7   from: 1   to: 466

       7 AAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGT 56
         :::|||||||:::      |||                |||:::   ::
      92 GlnIleLysArgAspArgAsn.................ProSerValIl 102

      57 CATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAAT 106
         :::::::   |||   :::   |||||||:::       :::    |||
     102 eValTrpIleGlyThrCysThrThrGluIleIleLysMetAspLeuGlu. 118

     107 ATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAA 156
                |||:::   |||:::::::::   :::   |||
     119 .....GlyLeuAlaProLysLeuGluAlaGluIleGlyIlePro...... 131

     157 AAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTAC 206
                                                  :::   ::
     132 .........................................IleValVa 134

     207 TGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCG 256
         :|||   |||:::|||          |||:::   :::||||||     |
     134 lAlaArgAlaAsnGlyLeuAspTyrAlaPheThrGlnGlyGluAspThrV 151

     257 TACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTC 306
         |||||   :::::::::
     151 alLeuAlaAlaMetAlaAlaArg........................... 158

     307 GAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCG 356
                      |||||||||   ::::::   ::::::::::::   ||
     159 .........CysProGluAlaAlaThrSerGluAlaAspGlnGlnGluAr 172

     357 CCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGT 406
          |   |||   ||||||                      |||       |
     172 gThrAsnAlaIleGln....................ArgLeuLeuGlnP 182

     407 TCAGCAAAATCCCTGCCGTAGATGCCGAG..................... 435
         ||:::|||   ||||||::::::   :::
     182 heGlyLysSerProAlaAlaGluGlnGlnProAlaSerSerLysHisPro 198

     436 CCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGA 485
         |||:::   :::|||   ::::::   |||      |||:::|||:::::
     199 ProLeuIleLeuPheGlySerValProAsp......ProValAlaThrGl 213

     486 CCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGT 535
         :   |||:::                  |||      |||:::|||:::
     213 nLeuThrIle..................GluLeuAlaLysGlnGlyIleT 224

     536 TGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCA 585
          |||   :::   |||         ::::::
     224 hrValSerGlyTrpLeuProAlaLysArgTyr.................. 234

     586 GGCGCAGACGTGCCG......TCTGAAAATGCTGCCAATATCGAAACACA 629
            ::::::::|||        :::|||::::::      |||      ::
     235 ...ThrGluLeuProValIleAlaGluGlySerTyrAlaIleGlyLeuAs 250

     630 TGAATTT........................................... 636
         :   |||
     250 nProPheLeuSerArgThrAlaThrThrLeuMetArgArgArgLysCysL 267

     637 ........GGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCAC...... 672
                 |||:::|||   |||   |||   :::|||::::::
     267 ysValIleGlyAlaProPheProIleGlyProAspGlySerArgAlaTrp 283

     673 ..................ATTCATTTCATCGAGCCAGTCGGC........ 696
                           :::      |||||||||   |||
     284 IleGluLysIleCysSerValLeuGluIleGluProGlnGlyLeuAlaGl 300

     697 .GCGAATAAAACCGTGTGGACCATC......AATTATCAAGACGTGATTG 739
            :::         ||||||         :::|||::::::::::::
     300 uArgGluAlaGlnValTrpAspSerIleGluAspTyrArgGlnLeuValG 317

     740 CTATCGGACGTTTGTTCGTAACAGGC...........CGTCTGAATACC 777
                :::::|||      |||             :::::::::
     317 luGlyLysGlnValPhePheMetGlyAspAsnLeuTrpGluIleSerLeu 333

     778 GAGCGCGTGGTTGCCTTGGGCGGCCTG.........CAAGTCAACAAACC 818
           |||   ::::::      |||:::         :::::::::   ||
     334 AlaArgPheLeuValArgCysGlyMetArgCysProGluIleGlyIlePr 350

     819 GCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCC. 867
         |    |||   |||      |||||||||::::::::::::|||    |||
     350 oTyrLeuAspArgArgTyrLeuGlyAlaGluLeuAlaMetLeuGluAlaT 367

     868 ........................GGCGAATTGGTTGAC...GCGGAC 888
                                  :::||||||:::         |||
     367 hrCysGlnSerMetGlyValProLeuProArgLeuValGluLysProAsp 383

     889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGC 938
         |||
     384 Asn............................................... 384

     939 GCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAG 988
                              :::|||||||||::::::   ||||||
     385 ......................TyrAsnGlnLeuGlnArgIleGlu.... 392

     989 GCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATAC 1038
                                   |||   |||||||||
     393 ............................AlaLeuGlnProAsp...... 397

    1039 TCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTC..... 1083
            ::::::   |||   ::::::|||          |||   |||
     398 ...LeuValIleThrGlyMetAlaHis......AlaAsnProLeuGluAl 411

    1084 ........................TTCAAGTTCACGACAGCCGTCAACG 1108
                                  :::||||||   |||
     411 aArgGlyIleSerThrLysTrpSerValGluPheThrPheAlaGlnIleH 428
```

```
1109 GCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCG 1158
         |||                       :::||||:::   :::   ::::::
 428 isGly.................PheGlyAsnAlaArgAlaIleLeuGlu 438

1159 TTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATAC 1208
      |||          |||       |||   |||:::|||   :::|||
 439 LeuValThrArgPro......LeuArgArgAsnLeuAlaLeuGlyThrLe 453

1209 CGACAGCGCGCAG 1221
        ::::::|||
 453 uGlyGlySerGln 457
```

seq_name: SwissProt_40:PTN8_MOUSE

```
seq_documentation_block:
ID   PTN8_MOUSE     STANDARD;      PRT;   802 AA.
AC   P29352;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Hematopoietic cell protein-tyrosine phosphatase 70Z-PEP (EC 3.1.3.48).
GN   PTPN8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92236615; PubMed=1373816;
RA   Matthews R.J., Bowne D.B., Flores E., Thomas M.L.;
RT   "Characterization of hematopoietic intracellular protein tyrosine
RT   phosphatases: description of a phosphatase containing an SH2 domain
RT   and another enriched in proline-, glutamic acid-, serine-, and
RT   threonine-rich sequences.";
RL   Mol. Cell. Biol. 12:2396-2405(1992).
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: SPLEEN, THYMUS, LYMPH NODE, AND BONE MARROW.
CC   -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC       TYROSINE PHOSPHATASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M90388; AAA39994.1; -.
DR   PIR; B44390; B44390.
DR   HSSP; Q06124; 2SHP.
DR   MGD; MGI:107170; Ptpn8.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_prot_phphtase.
DR   Pfam; PF00102; Y_phosphatase; 1.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00194; PTPc; 1.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW   Hydrolase.
FT   DOMAIN       23    288       PROTEIN-TYROSINE PHOSPHATASE.
FT   ACT_SITE    227    227       BY SIMILARITY.
SQ   SEQUENCE   802 AA;  89714 MW;  0F1E45339BD4613E CRC64;
```

```
alignment_scores:
           Quality:  112.00              Length:    351
             Ratio:  0.622                 Gaps:     20
Percent Similarity:  51.282    Percent Identity:  24.501
```

```
alignment_block:
 US-09-303-518D-131/rev x PTN8_MOUSE   ..

 Align seg 1/1   to: PTN8_MOUSE   from: 1   to: 802

 955 GTCCCAAATAATCAT...GCGCGCCTTGTGCAATCGCACCGTTCAATACC 909
     :::   :::||||||   :::|||    :::|||::::::   |||||||
 297 IleSerAspAsnHisLeuGlyArgGluIleGlnAlaGlnCysSerIlePr 313

 908 GAACCGG...........AAATCACGCGGTTGTCCGCGTCAACCAATTC 871
     |::::::            ::::::   :::||||||   :::      :
 313 oGluGlnSerLeuThrValGluAlaAspSerCysProLeuAspLeuProL 330

 870 GCCGG......CGGTAAGTTGAGACACCTTCG...CACCCAAAACGGTAC 830
     :::::        |||   |||:::|||   :::   |||   |||:::
 330 ysAsnAlaMetArgAspValLysThrThrAsnGlnHisSerLysGlnGly 346

 829 GCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGC 780
     ||| ||||||:::                         :::   |||:::   :::|
 347 Ala.GluAlaGlu.................SerThrGlyGlySerSerL 357

 779 TCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTC 730
     |||||::::::|||:::                 |||   ::::::::::::
 357 euGlyLeuArgThrSer...........ThrMetAsnAlaGluGluGlu 369

 729 TTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGA 680
     |||::::::   :::         :::               :::   |||   ||
 370 LeuValLeuHisSerAlaLysSerSerProSerPheAsnCysLeu...Gl 385

 679 AATGAATGTGCGTGCC.................ACTCAAGCCGGCAGGA 648
     |:::|||      :::                       |||:::   |||:::
 385 uLeuAsnCysGlyCysAsnAsnLysAlaValIleThrArgAsnGlyGlnA 402

 647 TGCGGGCCGCC......................AAATTCATGTGTTTC 622
      ||||||:::                          |||::::::   :::
 402 laArgAlaSerProValValGlyGluProLeuGlnLysTyrGlnSerLeu 418

 621 GATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACA 572
     |||:::||||||::::::   |||    |||:::|||:::||||||
 419 AspPheGlySerMetLeu.PheGly...SerCysProSerAlaLeuPro. 433

 571 CATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCG...... 528
         |||      :::   |||      |||::::::   |||
 434 .....IleAsnThrAlaAspArgTyrHisAsnSerLysGlyProValLys 448

 527 CGTTTGAAGTCTTCGGCGGCTTCTTTGATG.................... 498
     |||    |||||||:::      :::|||:::
 449 ArgThrLysSerThrProPheGluLeuIleGlnGlnArgLysThrAsnAs 465

 497 .ATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCC............ 462
      ::::::||||||::::::   |||   |||   |||
 465 pLeuAlaValGlyAspGlyPheSerCysLeuGluSerGlnLeuHisGluH 482

 461 ................ATCGCATTGACGAAGATGGCGAACGGCTCGGCA 429
                    :::   :::   ::::::|||:::   |||:::
 482 isTyrSerLeuArgGluLeuGlnValGlnArgValAlaHisValSerSer 498

 428 TCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTC..... 384
     :::      :::         |||   |||         ::::::
 499 GluGluLeuAsnTyrSerLeuProGlyAlaCysAspAlaSerCysValPr 515

 383 ....CATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCA 338
        |||:::|||      :::|||::::::         ||||||    :
 515 oArgHisSerProGlyAlaLeuArgValHisLeuTyrThrSerLeuAlaG 532

 337 ATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTG 288
     ::               |||                        ||||||
 532 luAspProTyrPheSer.......................SerSerPro 540

 287 CCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTG 238
     |||::::::|||         |||   :::      |||:::
 541 ProAsnSerAla.........AspSerLysMetSerPheAsp........ 551
```

```
237 AATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGAT 188
                 |||||||||               |||:::||||||
552 .............LeuProGluLysGlnAspGlyAlaThrSerProGlyA 564

187 TCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTGACGGCTTCACCTTCC 138
     :::|||   ::::::||||||       |||:::   ::::::|||
564 laLeuLeuProAlaSerSerThrThrSerPhePheTyrSerAsnProHis 580

137 TTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGC 88
       ::::::   :::   :::   ::::::|||   |||   |||
581 AspSerLeuValMetAsnThrLeuThrSerPheSerProProLeuAsnGl 597

 87 GACTTCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTGCCC 39
       :::::::::   :::||||||       |||   :::   :::|||
597 nGluThrAlaValGluAlaProSerArgArgThrAspAspGluIlePro 613
```

seq_name: SwissProt_40:PPSA_PYRFU

seq_documentation_block:

```
ID   PPSA_PYRFU     STANDARD;      PRT;   817 AA.
AC   P42850; Q59672;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate,water
DE   dikinase) (PEP synthase).
GN   PPSA OR PF0043.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=2261;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vc1 / DMS 3638;
RX   MEDLINE=95129854; PubMed=7828869;
RA   Robinson K.A., Schreier H.J.;
RT   "Isolation, sequence and characterization of the maltose-regulated
RT   mlrA gene from the hyperthermophilic archaeum Pyrococcus furiosus.";
RL   Gene 151:173-176(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vc1 / DMS 3638;
RX   MEDLINE=95354939; PubMed=7628701;
RA   Jones C.E., Fleming T.M., Piper P.W., Littlechild J.A., Cowan D.A.;
RT   "Cloning and sequencing of a gene from the archaeon Pyrococcus
RT   furiosus with high homology to a gene encoding phosphoenolpyruvate
RT   synthetase from Escherichia coli.";
RL   Gene 160:101-103(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC       phosphoenolpyruvate + phosphate.
CC   -!- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC       LACTATE ARE USED AS A CARBON SOURCE.
CC   -!- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U08376; AAA81512.1; -.
DR   EMBL; X80819; CAA56785.1; -.
DR   HSSP; P22983; 1DIK.
DR   InterPro; IPR000121; PEP_utilizers.
DR   InterPro; IPR002192; PPDK_N_term.
DR   Pfam; PF00391; PEP-utilizers; 1.
DR   Pfam; PF02896; PEP-utilizers_C; 1.
DR   Pfam; PF01326; PPDK_N_term; 1.
DR   ProDom; PD000940; PEP_utilizers; 1.
DR   PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR   PROSITE; PS00742; PEP_ENZYMES_2; 1.
KW   Transferase; Kinase; ATP-binding; Phosphorylation.
FT   MOD_RES      442    442       PHOSPHORYLATION (BY SIMILARITY).
FT   DOMAIN       809    815       POLY-GLU.
FT   CONFLICT     747    747       Q -> K (IN REF. 2).
SQ   SEQUENCE   817 AA;  90485 MW;  582694CEF13C74BA CRC64;
```

alignment_scores:

```
          Quality:  111.00           Length:    489
            Ratio:  0.459              Gaps:     24
Percent Similarity:  49.489   Percent Identity: 21.268
```

alignment_block:

```
US-09-303-518D-131 x PPSA_PYRFU  ..

Align seg 1/1  to: PPSA_PYRFU  from: 1  to: 817

 31 CCCATCGCGGGCAGACCGGAGCAAGTCATTTATGAC.............. 66
    |||::::::::::::   ::::::::::::   :::
235 ProValThrAsnAsnArgAsnGluIleMetIleAsnAlaSerTrpGlyLe 251

 67 .GGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCA 115
      |||   |||:::::::::   |||:::         :::|||||||:::
251 uGlyGluAlaValValSerGlyAlaValThrProAspGluTyrIleValG 268

116 TGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAA 165
     :::   :::   ||||||||||||         :::||||||
268 luLysGlyThrTrpLysIleLysGluLysValIleAlaLysLysGluVal 284

166 GTGCTGTTTGAAGACAAAAAGAATCCG............GGCGTAGTATT 203
    :::::::::         :::||||||            |||:::|||::
285 MetValIle.........ArgAsnProGluThrGlyArgGlyThrValMe 298

204 TACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGC....... 246
    ::::                      |||:::|||   :::   |||
298 tVal..............LysValAlaGlu...TyrLeuGlyProGluT 309

247 .....GAAAAGCGCGTACTT......CAGTCAGTCGTGATTGCCGTTGAA 285
         ||||||:::||||||       ::::::::::::   :::::::::
309 rpValGluLysGlnValLeuThrAspGluGlnIleIleGluValAlaLys 325

286 GGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCA.. 333
      :::::::::||||||   :::         |||:::   :::
326 MetGlyGlnLysIleGluAspHisTyrGlyTrpProGlnAspIleGluTr 342

334 .AAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTAT 382
       :::::::::      :::      ::::::||||||
342 pAlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerArgProI 359

383 GGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCC 432
      |||::::|||:::   :::            |||   ::::::
359 leThrThrLeuLysGluGluAlaThrAlaGlu...GluAlaGluGluVal 374

433 GAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG...CTGGC 479
    |||         ::::::::::::::::::   ::::::|||   :::::
375 GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyIleGl 391

480 TGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAA......GACTTCA 523
    :|||       :::|||:::|||   :::|||:::|||           |
391 yAlaGlyArgValValValIlePheAspAlaSerGluIleAspLysValL 408

524 AACGCGGC...CTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCAT 570
    ||:::|||   :::||||||   :::       :::         :::
408 ysGluGlyAspIleLeuValThrThrMetThrAsnProAspMetValPro 424

571 GTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCT........ 612
    :::   |||   |||:::|||   |||   :::|||::::::
425 AlaMetLysArgAlaAlaAlaIleValThrAspGluGlyGlyArgThrSe 441

613 ....GCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCT 658
        |||   |||   ::::::|||:::|||   |||
441 rHisAlaAlaIleValSerArgGluLeuGlyIlePro............. 453

659 TGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACC 708
```

```
 453 ................................................... 453

 709 GTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGT 758
                                  ::::::|||       ::
 454 ..........................CysValValGlyThrLysGluAl 461

 759 AACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAA. 807
     :|||   :::|||:::|||           |||::::::  |||   :::
 461 aThrLysLysLeuLysThrGlyMetTyrValThrValAspGlyThrArgG 478

 808 .....GTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTG 852
          |||   |||     :::          |||      :::
 478 lyLeuValTyrLysGlyIleValLysSerLeuValLysLysLysGluGlu 494

 853 TCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTC 902
     ::::::       :::|||::::::|||   |||         ::::::::
 495 AlaLysAlaGluGlyGlyGlnValValValAlaGlyAlaProLeuValTh 511

 903 CGGTTCG............................GTATTGAACGGTG 922
     :|||:::                            |||   :::.  |
 511 rGlyThrMetValLysValAsnValSerMetProGluValAlaGluArgA 528

 923 CGATTGCACAAGGCGCG..........................CATGAT 945
      ||   |||   ||||||                          |||
 528 laAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAlaGluHisMet 544

 946 TATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAG 995
         |||:::       ::::::    |||:::    |||:::||||||::::::
 545 IleLeuSerIleGlyGlnHisProIleLysPheIleLysGluGlyLysGl 561

 996 CAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCA 1045
     ::::||||||         :::|||          :::|||    :::
 561 uGluGluLeuValGluLysLeuAlaGluGlyIleGluLysValAlaAla. 577

1046 CGCGCACCACTCTCGGCCATTT....................CCTAAAA 1074
       |||    :::   ||||||                       |||    :
 578 .AlaPheTyrProArgProValTrpTyrArgThrLeuAspAlaProThrA 594

1075 AACAAACTCTTCAAGTT.........CACGACAGCCGTCAACGGCGGCGA 1115
     ::::      ::::::           :::|||:::   ::::::|||:::
 594 snGluPheArgGluMetProGlyGlyGluAspGluProGluGluArgAsn 610

1116 CCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACA 1165
     |||         |||                     |||:::::::::    ::
 611 ProMetLeuGlyTrpArgGly.........IleArgArgGlyLeuAspGl 624

1166 TCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGC 1215
     :|||        |||        |||:::|||:::       :::         :
 624 nProGluLeuLeuArg...AlaGluPheLysAlaIleLysLysValValG 640

1216 GCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTG 1265
     ::    |||::::::    :::|||                        |||
 640 luLysGlyTyrAsnAsnIleGly.......................Val 648

1266 CAGCTTCGTCTG.........CCCGGGCAAATACGAATACGGCCCGCTGT 1306
     :::|||    |||            |||    |||||||||           ::
 649 MetLeuProLeuValSerHisProGluGlnIleArgGluAlaLysArgIl 665

1307 TGCGCAAAGTGCTGGA 1322
     :|||::::::::::|||
 665 eAlaArgGluValGly 670
```

seq_name: SwissProt_40:HFC1_MESAU

seq_documentation_block:

```
ID   HFC1_MESAU     STANDARD;      PRT;  2090 AA.
AC   P51611;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF)
DE   (CFF).
GN   HCFC1.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goto H., Nishitahi H., Umene K.I., Nakabepu Y., Nishimoto T.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC       TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC       ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC       PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC       ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC       110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N-
CC       AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC       NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC       SIGNAL (BY SIMILARITY).
CC   -!- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC       CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC       AT A DEFINED SITE, PPCE/THET, WITHIN THE HCF REPEAT (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D45419; BAA08258.1; -.
DR   HSSP; P02751; 1FNA.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001798; Kelch.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF01344; Kelch; 4.
DR   SMART; SM00060; FN3; 2.
KW   Nuclear protein; Repeat.
FT   REPEAT       44     89       KELCH 1.
FT   REPEAT       93    140       KELCH 2.
FT   REPEAT      148    194       KELCH 3.
FT   REPEAT      217    265       KELCH 4.
FT   REPEAT      266    313       KELCH 5.
FT   DOMAIN     1010   1448       8 X 26 AA APPROXIMATE REPEATS.
FT   REPEAT     1010   1035       HCF REPEAT 1.
FT   REPEAT     1072   1097       HCF REPEAT 2.
FT   REPEAT     1101   1126       HCF REPEAT 3.
FT   REPEAT     1157   1182       HCF REPEAT 4 (DEGENERATE).
FT   REPEAT     1295   1320       HCF REPEAT 5.
FT   REPEAT     1323   1348       HCF REPEAT 6.
FT   REPEAT     1358   1383       HCF REPEAT 7 (DEGENERATE).
FT   REPEAT     1423   1448       HCF REPEAT 8.
SQ   SEQUENCE   2090 AA;  214942 MW;  E495E8B1F2385E17 CRC64;
```

alignment_scores:

```
          Quality:  109.50            Length:    478
            Ratio:  0.534               Gaps:     22
 Percent Similarity:  42.887   Percent Identity: 21.339
```

alignment_block:

```
 US-09-303-518D-131/rev x HFC1_MESAU   ..

 Align seg 1/1  to: HFC1_MESAU  from: 1  to: 2090

    1118 CGGTCGCCGCCGTTGACGGCTGTC...........GTGAACTTGAAGAG 1081
```

```
                 :::  ||||||  :::  |||           :::|||  :::||
 363 LysProProProProAlaArgValGlnLeuValArgAlaAsnThrAsnSe 379

1080 TTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTTG. 1032
     ||||        :::|||     |||:::  :::  :::||||||
 379 rLeuGluVal...SerTrpGlyAlaValAlaThrAlaAspSerTyrLeuL 395

1031 ............................TCCGGCTGCGGCGCAACCCAG 1011
                                 :::::::::  |||||||:::
 395 euGlnLeuGlnLysTyrAspIleProAlaThrAlaAlaThrAlaThrSer 411

1010 CCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTG 961
     |||:::                |||:::  :::  ::::::
 412 ProThr..............ProAsnProValProSerValProAla.. 422

 960 GTAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCG....... 918
        :::|||   |||  ||||||:::|||  ||||||
 423 ....AsnProProLysSerProAlaProAlaAlaAlaAlaProAlaValG 438

 917 ...................TTCAATACCGAACCGGAAATCACGCGGTTG 888
                        ::::::     |||:::  :::
 438 lnProLeuThrGlnValGlyIleThrLeuValProGlnAlaAlaAlaAla 454

 887 TCCGCGTCAACCAATTCG............CCGGCGGTAAGTTGAGACAC 850
         ||||||::::::             |||:::|||      :::::
 455 ProProSerThrThrThrIleGlnValLeuProThrValProGlySerSe 471

 849 CTTCGCACCCAAAACGGTACGCAAGAGGCGCGGT.........TTGTTGA 809
     ::::::      |||:::  :::  :::|||            :::|||
 471 rIleSerValProThrAlaAlaArgAlaGlnGlyValProAlaValLeuL 488

 808 CTTGCAGGCCGCCCAAGGCAACCACGCGCTCG......GTATTCAGACGG 765
               |||:::|||||||||  :::      |||    |||
 488 ysValThrGlyProGlnAlaThrThrGlyThrProLeuValThrMetArg 504

 764 CCTGTT......ACGAACAAACGTCCGATAGCAATCACGTCTTGATAATT 721
     |||:::      :::::::|||  |||:::::::::||||||     ||
 505 ProAlaGlyGlnAlaGlyLysAlaProValThrValThrSer......Le 519

 720 GATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGT 671
     |  :::  :::  :::  :::|||||||  |||  :::
 519 uProAlaSerValArgMetValValProThrGlnSerAlaGlnGlyThrV 536

 670 GCGTGCCACTCAAGCCGGCAGGATGCGGG..................... 642
        :::    :::|||     |||
 536 alIleGlySerAsnProGlnMetSerGlyMetAlaAlaLeuAlaAlaAla 552

 641 ....................CCGCCAAATTCATGTGTTTCGATATTGGC 613
                          ||||||:::|||:::  :::::|||::
 553 AlaAlaAlaThrGlnLysIleProProSerSerAlaProThrValLeuSe 569

 612 A................................GCATTTTCAGACGGCA 596
     :                                |||  :::  ||||
 569 rValProAlaGlyThrThrIleValLysThrValAlaValThrProGlyT 586

 595 CGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGG 546
     ||::::::    |||                    |||
 586 hrThrThrLeuProAlaThrValLysValAlaSerSerProValMetVal 602

 545 CTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGAT 496
        |||          ||||||:::::||||||::::::
 603 SerAsnProAlaThrArgMetLeuLysThrAlaAlaAlaGlnValGlyTh 619

 495 GACCGTAGGGTCGGCAGCC............................... 477
      :::|||:::|||||||||
 619 rSerValSerSerAlaAlaAsnThrSerThrArgProIleIleThrValH 636

 476 .....AGCGGATTGGTGTCCATCGCA........................ 456
          ||||||  |||::::::|||
 636 isLysSerGlyThrValThrValAlaGlnGlnAlaGlnValValThrThr 652

 456 .................................................. 456

 653 ValValGlyGlyValThrLysThrIleThrLeuValLysSerProIleSe 669

 456 .................................................. 456

 669 rValProGlyGlySerAlaLeuIleSerAsnLeuGlyLysValMetSerV 686

 455 ...................TTGACGAAGATGGCGAACGGCTCGGCATCT 426
                         |||:::  ::::::|||:::||||||
 686 alValGlnThrLysProValGlnThrSerAlaValThrGlyGlnAlaSer 702

 425 ACG.....................................GCAGG 418
     |||                                      |||||
 703 ThrGlyProValThrGlnIleIleGlnThrLysGlyProLeuProAlaGl 719

 417 GATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAA 368
     |  :::|||:::        |||  ||||||    ||||||
 719 yThrIleLeuLys...LeuValThrSerAlaAspGlyLysProThr.... 733

 367 TCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGCTTCAGGT 318
                          ::::::::::::  ::::::|||||||||
 734 ...................ThrIleIleThrThrThrGlnAlaSerGly 743

 317 ACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCAACGGCAATCACGAC 268
     :::     :::  :::|||  ::::::  |||::::::  :::||
 744 AlaGlySerLysProThrIleLeuGlyIleSerSerValSerProSerTh 760

 267 T.........GACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCGA 227
     |               :::|||  ::::::          ::::::||||
 760 rThrLysProGlyThrThrThrIleIleLysThrIleProMetSerAlaI 777

 226 TTTTGCCTGAAGCCGGCGCA......GTAAATACTACGCCCGGATTC... 186
     |||:::  :::|||||||||      |||:::::::|||||||||:::
 777 leIleThrGlnAlaGlyAlaThrGlyValThrSerThrProGlyIleLys 793

 185 ..........................TTTTGTCTTCAAACAGCACTTG 163
                               ::::::|||::::::
 794 SerProIleThrIleIleThrThrLysValMetThrSerGlyThrGlyAl 810

 162 GCCT......TTTTGACGGCTTCACCTTCCTTG 135
       |||       ::::::||||||   |||::::::
 810 aProAlaLysIleIleThrAlaValProLysIle 821
```

seq_name: SwissProt_40:N121_HUMAN

```
seq_documentation_block:
ID   N121_HUMAN      STANDARD;      PRT;  1229 AA.
AC   Q9Y2N3; O75115; Q9Y4S7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE   of 121 kDa) (P145).
GN   NUP121 OR KIAA0618.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cordes M., Bauer C., Holmes A.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 243-1229 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [3]
RP   SEQUENCE OF 1130-1229 FROM N.A.
```

```
RC   TISSUE=Uterus;
RA   Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE
CC       REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS
CC       OF THE PORE COMPLEX TO THE PORE MEMBRANE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC       MEMBRANE (BY SIMILARITY).
CC   -!- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC   -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC       SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC       NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AC006014; AAD28064.1; -.
DR   EMBL; AB014518; BAA31593.1; -.
DR   EMBL; AL080109; CAB45713.1; -.
KW   Nuclear protein; Transport; Transmembrane; Repeat.
FT   DOMAIN        1     40       CISTERNAL SIDE (POTENTIAL).
FT   TRANSMEM     41     61       POTENTIAL.
FT   DOMAIN       62   1229       PORE SIDE (POTENTIAL).
FT   DOMAIN        4     10       POLY-ALA.
FT   DOMAIN       51     56       POLY-ALA.
FT   DOMAIN      294    299       POLY-LYS.
FT   DOMAIN      441    444       POLY-SER.
FT   DOMAIN      499    502       POLY-PRO.
FT   DOMAIN      733    736       POLY-SER.
FT   DOMAIN      819    826       POLY-SER.
FT   DOMAIN      869    875       POLY-THR.
FT   DOMAIN     1061   1067       POLY-SER.
FT   CONFLICT    297    297       K -> E (IN REF. 2).
FT   CONFLICT    336    336       S -> N (IN REF. 2).
FT   CONFLICT    379    379       T -> A (IN REF. 2).
FT   CONFLICT    451    451       K -> R (IN REF. 2).
FT   CONFLICT    545    545       T -> I (IN REF. 2).
FT   CONFLICT    559    559       T -> P (IN REF. 2).
FT   CONFLICT    702    702       S -> P (IN REF. 2).
FT   CONFLICT    881    881       P -> H (IN REF. 2).
FT   CONFLICT    965    965       T -> A (IN REF. 2).
FT   CONFLICT    982    984       PAT -> A (IN REF. 2).
FT   CONFLICT    991    996       ASTIKI -> PSMIKV (IN REF. 2).
FT   CONFLICT   1000   1000       H -> Y (IN REF. 2).
FT   CONFLICT   1006   1008       QPT -> HPI (IN REF. 2).
FT   CONFLICT   1044   1044       MISSING (IN REF. 2).
FT   CONFLICT   1102   1102       A -> T (IN REF. 2).
FT   CONFLICT   1124   1124       T -> A (IN REF. 2).
FT   CONFLICT   1146   1146       S -> G (IN REF. 1).
FT   CONFLICT   1165   1165       L -> Q (IN REF. 1).
FT   CONFLICT   1195   1195       A -> G (IN REF. 1).
FT   CONFLICT   1199   1229       SAAPSFSIGAGSKTPGARQRLQARRQHTRKK -> NTFAHQ
FT                                QEHSPRKGPNNLSKRKLLPAVRAQGPPRRGQASSFPTRKE
FT                                (IN REF. 2).
FT   CONFLICT   1202   1202       P -> L (IN REF. 3).
SQ   SEQUENCE   1229 AA;  125087 MW;  515655D12858998B CRC64;

alignment_scores:
           Quality:  108.50          Length:     524
             Ratio:  0.478             Gaps:      22
 Percent Similarity:  43.321   Percent Identity: 21.565

alignment_block:
 US-09-303-518D-131/rev x N121_HUMAN   ..

 Align seg 1/1  to: N121_HUMAN  from: 1  to: 1229

1310 CGCAACAGCGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAA 1261
     :::||||||:::        |||       ||||||:::::       :::||
 467 GlnAsnSerAsnSerGlnSer...ThrProGlySerSerGlyGlnArgLy 482

1260 AGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGT 1211
     |   :::   :::          |||:::::   :::      :::|||
 482 sArgLysValGlnLeuLeuProSerArgArgGlyGluGlnLeuThrLeuP 499

1210 CGGTATCGCCGACGATT................................. 1194
                |||   :::
 499 roProProProGlnLeuGlyTyrSerIleThrAlaGluAspLeuAspLeu 515

1193 ..............................AAATCGCGCAAAAGCAAGGT 1174
                                   ::::::   ::::::|||
 516 GluLysLysAlaSerLeuGlnTrpPheAsnGlnAlaLeuGluAspLysSe 532

1173 AGGCAGGATGTCCAACGGCATTACGCGCTCA................... 1143
                ||||||:::::|||::::::
 532 rAspAlaAlaSerAsnSerValThrGluThrProProThrThrGlnProS 549

1142 ..........TAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTG 1104
                  :::|||   ::::::   |||    ||||||||
 549 erPheThrPheThrLeuProAlaAlaAlaThrAla...SerProProThr 564

1103 ACGGCTGTCGTGAAC...............TTGAAGAGTTTGTTTTTTAG 1069
     :::    ::::::                  |||:::||||||
 565 SerLeuLeuAlaProSerThrAsnProLeuLeuGluSerLeuLysLysMe 581

1068 GAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTTGTCCGGCTGCGGCG 1019
      :::   |||       :::                :::|||::::::
 581 tGlnThrProProSerLeuProProCysProGluSerAlaGlyAlaAlaT 598

1018 CAACCCAGCCGAACAGC...............TCTTTGCTGCGGCCTTCT 984
     ::|||:::      |||               ||||||||    |||
 598 hrThrGluAlaLeuSerProProLysThrProSerLeuLeuProProLeu 614

 983 TCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCC 934
     ::::::::::::                                 :::||
 615 GlyLeuSerGln.............................SerGlyPr 621

 933 TTGTGCAATCGCACCG.........TTCAATACCGAACCGGAAATCACGC 893
     |   ::::::   |||          |||:::::::|||        |||
 621 oProGlyLeuLeuProSerProSerPheAspSerLysProProThrThrL 638

 892 GGTTGTCC.........GCGTCAACCAATTCGCCGGCGGTAAGTTGAGAC 852
       |||:::         |||   :::      ||||||:::      |||
 638 euLeuGlyLeuIleProAlaProSerMetValProAlaThr......Asp 652

 851 ACCTTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAG 802
     |||   ||||||   |||:::::   :::            |||:::
 653 ThrLysAlaProProThrLeuGlnAlaGlu............ThrAlaTh 665

 801 GCCGCCCAAGGCAACCACG.................CGCTCGGTATTCA 770
          |||:::||||||:::                  ::::::   |||
 665 rLysProGlnAlaThrSerAlaProSerProAlaProLysGlnSerPheL 682

 769 GACGGCCTGTTACGAACAAACGTCCGATA...............GCAATC 735
              :::   |||      |||                 |||
 682 euPheGlyThrGlnAsnThrSerProSerSerProAlaAlaProAlaAla 698

 734 ACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTC 685
      :::|||         |||       ::::::   ||||||      ||
 699 SerSerAlaSerProMetPheLysProIlePheThrAlaProProLysSe 715

 684 GATGAAATGAATGTGCGTGCCACTCAAGCCG................... 654
     |   |||          :::|||       |||
 715 rGluLysGluGlyProThrProProGlyProSerValThrAlaThrAlaP 732

 654 .................................................. 654

 732 roSerSerSerSerLeuProThrThrThrSerThrThrAlaProThrPhe 748

 653 ...........GCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATT 616
```

```
                  ::::::  |||||||||  |||  |||  :::
749 GlnProValPheSerSerMetGlyProProAlaSer...ValProLeuPr 764

615 GGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGA 566
     |||  |||       |||::::::|||||||:::     |||
764 oAlaProPhePheLysGlnThrThrThrProAlaThrAlaProThr.... 779

565 TTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAG... 519
                                 |||:::  |||  :::
780 .........................ThrThrAlaProLeuPheThrGly 787

518 ...TCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAGCGG 472
      :::|||||||:::|||  :::     :::  ||||||:::  ::
788 LeuAlaSerAlaThrSerAlaValAlaProIleThrSerAlaSerProSe 804

471 ATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGC......TCGGCAT 428
     :     |||  :::        ::::::  |||      |||:::|
804 rThrAspSerAlaSerLysProAlaPheGlyPheGlyIleAsnSerValS 821

427 CTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAG 378
    ||::::::::::::      :::      :::   :::|||:::  :::
821 erSerSerSerValSerThrThrThrSerThrAlaThrAlaAlaSerGln 837

377 CCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAG 328
    |||                        |||:::      :::   :::::
838 Pro.................PheLeuPheGlyAlaProGlnAlaSerAl 848

327 CGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTG......CCTT 284
     :||||||  |||        :::|||||||  ::::::     |||
848 aAlaSerPheThrProAlaMetGlySerIlePheGlnPheGlyLysProP 865

283 CAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATA 234
      :::      ||||||          |||  ||||||
865 roAlaLeuProThrThrThrThrValThrThrPheSerGlnSer...... 879

233 GCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACG........... 195
                ||||||  |||      :::  |||:::
880 .........LeuProThrAlaValProThrAlaThrSerSerSerAlaAl 893

194 .......CCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGA 152
                ||||||      :::      ::::::  |||      :
893 aAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAlaThrS 910

151 CGGCTTCACCTTCCTTGATTTCATCGAGGGGCGCATGCCGACATATTCT 102
     ::::::::|||:::|||  |||   :::         ||||||::::::
910 erSerGlnProThrLeuThrPheSerAsnThrSerThrProThrPheAsn 926

101 TCGCCAAGCAACGCGACTTCGGTAATGGCCGGGCCGTCATAAATGACTTG 52
       |||   ::::::::::::              ||||||       ::
927 IleProPheGlySerSerAlaLysSerProLeuProSerTyrProGlyAl 943

 51 CTCCGGTCTGCCCGCGATGGGC 30
     ::::     |||||||:::|||
943 aAsnProGlnProAlaPheGly 950
```

seq_name: SwissProt_40:MSB2_YEAST

```
seq_documentation_block:
ID   MSB2_YEAST     STANDARD;      PRT;  1306 AA.
AC   P32334;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MSB2 protein (Multicopy suppression of a budding defect 2).
GN   MSB2 OR YGR014W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92383951; PubMed=1514328;
RA   Bender A., Pringle J.R.;
RT   "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RT   defect.";
RL   Yeast 8:315-323(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=97435481; PubMed=9290212;
RA   Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT   "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT   chromosome VII.";
RL   Yeast 13:1077-1090(1997).
CC   -!- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC   -!- SIMILARITY: SOME, TO YEAST HKR1.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M77354; AAA34798.1; -.
DR   EMBL; Z72799; CAA96997.1; -.
DR   PIR; S25370; S25370.
DR   SGD; S0003246; MSB2.
KW   Transmembrane; Glycoprotein; Repeat.
FT   DOMAIN      698    816       7 X 17 AA TANDEM REPEATS.
FT   REPEAT      698    714       1.
FT   REPEAT      715    731       2.
FT   REPEAT      732    748       3.
FT   REPEAT      749    765       4.
FT   REPEAT      766    782       5.
FT   REPEAT      783    799       6.
FT   REPEAT      800    816       7.
SQ   SEQUENCE   1306 AA;  133114 MW;  67D5D984D5CA4A6D CRC64;

alignment_scores:
          Quality:  108.00            Length:    512
            Ratio:  0.468               Gaps:     21
 Percent Similarity:  45.117   Percent Identity:  21.875

alignment_block:
 US-09-303-518D-131/rev x MSB2_YEAST  ..

 Align seg 1/1  to: MSB2_YEAST  from: 1  to: 1306

 1342 AGCCTTCCTTCTCAATGGTT.......TCCAGCACTTTGCGCAACAGCGG 1300
      ||||||||||||  |||        |||||||||   :::::::|||
   83 SerLeuProSerThrTrpTyrAspAlaSerSerThrSerGlnThrSerVa 99

 1299 GCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTT 1250
         |||:::          ::::::      :::::::::          |
   99 lSerTyrAlaSerGlnGluSerAspTyrAlaValAsnGlnAsnSerTrpS 116

 1249 CTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCG...GTATCG 1203
      ||:::|||:::::::::   |||:::::::   :::   |||      :::
  116 erAlaSerThrAsnGlnLeuProSerThrSerThrThrSerTyrTyrAla 132

 1202 CCGACGATTAAATCG...........CGCAAAAGCAAGGTAGGCAGGAT 1165
      ||||||:::::::::                     |||:::|||:::
  133 ProThrPheSerThrSerAlaAspPheAlaAlaSerSerValAsnAlaAl 149

 1164 GTCCAACGGCATTACGCGCTCATAAGTGCCGATCGGTACCATGGCGCGGT 1115
      |||:::         |||   |||   |||||||||   |||   |||:::|
  149 aSerAspValSerThrAlaSer...ValProIleAspThrSerAlaAsnS 165

 1114 CGCCGCCGTTGACGGCTGTCGTGAACTTGAAG................. 1083
      ||    |||:::|||::::::    |||::::::
  165 erIleProPheThrThrThrSerAsnIleGluThrThrThrSerAlaPro 181

 1083 .................................................. 1083
```

```
 182 LeuThrSerAspThrProLeuIleSerThrSerThrMetSerAlaAlaAs 198

1082 .AGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTG.......... 1044
      :::::|||          |||      :::   :::
 198 pAsnValPheSerSerAlaAsnProIleSerAlaSerLeuThrThrThrA 215

1043 ..............ATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCG 1008
                   ::::::     |||   :::||||||
 215 spSerSerGluSerPheAspGlnThrSerThrAlaGlyAlaIleProVal 231

1007 AACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTA 958
      :::||||||
 232 GlnSerSer......................................... 234

 957 GCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCG 908
                                      |||   |||:::::::
 235 ..................................AlaAspPheSerSerS 240

 907 AACCGGAAATC...ACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTA 861
      ::   ||||||   ::::::   |||||||:::   :::|||||||:::
 240 erSerGluIleLeuValGlnSerSerAlaAspPheSerSerProSer... 255

 860 AGTTGAGACACCTTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTT 811
                      :::|||   |||:::            :::
 256 ...............SerProThrThrThr...........AspIleSe 263

 810 GACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTG 761
                   |||      |||::::::|||               :
 263 rLeuSerAlaAlaProLeuGlnThrSerGluSerSerSerPheThrThrA 280

 760 TTACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCAC 711
      :::::        |||:::::::   |||:::
 280 laSerAlaAlaLeuProValSerSerThrAsp.................. 290

 710 ACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCGTGCCACT 661
                           ||||||               |||   ::
 291 ...............ValAspGlySerSerAlaSerProValValSerMe 302

 660 CAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAG 611
      ::::   ||||||                                  |
 302 tSerAlaAlaGly.............................GlnIleA 309

 610 CATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTA 561
      ||   |||::::::|||:::   |||:::      :::|||:::   |||
 309 laSerSerSerSerThrAspAsnProThrMetSerGluThrPheSerLeu 325

 560 CGTTCGGTCAGG.............................CGGCTCAA 541
        |||::::::
 326 ThrSerThrGluValAspGlySerAspValSerSerThrValSerAlaLe 342

 540 TACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCG 491
         :::   |||   |||::::::|||:::::::::   :::      |
 342 uLeuSerAlaProPheLeuGlnThrSerThrSerAsnSerPheSerIleV 359

 490 TAGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCG 441
      ||:::   ::::::|||   :::   |||   :::   :::   :::|||
 359 alSerProSerValSerPheValProSerGlnSerSerSerAspValAla 375

 440 AACGGCTCGGCATCTACGGCAGGGATTTTG.................... 411
           |||:::|||||||||:::::::::
 376 ......SerSerSerThrAlaAsnValValSerSerSerPheSerAspIl 390

 410 ................CTGAACGGACGGGTGCGAAGCGCAGTCCATAAGC 377
                       :::|||   |||
 390 eProProGlnThrSerThrSerGlySerVal................... 400

 376 CTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGC 327
                :::   :::   ::::::   |||   |||   |||   |||
 401 ........ValSerValAlaGlnSerAlaSerAlaLeuAlaPheGlnSer 414

 326 GCTTCA.........GGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCC 286
      ::::::         |||:::         |||:::::::::||||||||

 415 SerThrGluValTyrGlyAlaSerAlaSerSerThrMetSerSerLeuLe 431

 285 TTCAACG...........GCAATCACGACTGACTGAAGTACGCGCTTTT 248
      ||||||            :::   |||   |||   |||:::      |
 431 uSerThrThrSerLeuGlnSerThrThrLeuAspSerSerSerLeuAlaS 448

 247 CGCCACGGTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACT 198
      ||            ::::::   |||   :::   |||:::   :::|||
 448 erSerSerAlaSerSerSerAspLeuThrAspTyrGlyValSerSerThr 464

 197 ACGCCCGGATTCTTTTTGTCT...........TCAAACAGCACTTGGCC 160
      :::         :::||||||            :::::::::::
 465 AlaSerIleProLeuLeuSerAlaSerGluGlnAlaSerThrSerSerSe 481

 159 TTTTTGACGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGC....... 117
       |||   ::::::|||||||||:::   |||:::   ::::::
 481 rPheSerValValSerProSerValSerPheValProSerGlnSerSerS 498

 116 .....ATGCCGACATATTCTTCGCCAAGC.................AAC 90
           :::   :::   |||:::||||||                 :::
 498 erAspValAlaSerThrSerAlaProSerValValSerSerSerPheSer 514

  89 GCGACTTCGGTAATGGCCGGGCCGTCATAAATGACT 54
        ||||||::::::||||||   |||   ||||||
 515 TyrThrSerLeuGlnAlaGlyGlySerSerMetThr 526
```

seq_name: SwissProt_40:AMYH_YEAST

```
seq_documentation_block:
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE   glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC       glucose residues successively from non-reducing ends of the chains
CC       with release of beta-D-glucose.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z38061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA32069.1; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; S0001458; MUC1.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN      210   1367       SER/THR-RICH.
FT   CARBOHYD    817    817       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;

alignment_scores:
          Quality:  107.00          Length:    546
            Ratio:  0.473             Gaps:     19
 Percent Similarity:  41.392   Percent Identity: 19.963

alignment_block:
 US-09-303-518D-131/rev x AMYH_YEAST   ..

 Align seg 1/1  to: AMYH_YEAST  from: 1  to: 1367

   1337 TCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTCGTA 1288
        |||||||||      :::::::::       :::|||:::   |||
    891 SerPheSerThrGlyThrThrValThrProSerSerSerLysTyr..... 905

   1287 TTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATT 1238
            |||||||:::              ::::::  |||:::::::::::::
    906 ....ProGlySerGlnThrGluThrSerValSerSerThrThrGluThrT 921

   1237 CCAAGCAACCCAAAGCCTGCGCGCTGTCGGTA............TCGCCG 1200
         ::       |||        :::   ||||||          :::
    921 hrIleValProThrLysThrThrThrSerValThrThrProSerThrThr 937

   1199 ACGATTAAATCGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTAC 1150
        ||||||  :::        ::::::|||     |||   |||   ||
    938 ThrIleThrThrThrValCysSerThrGlyThrAsnSerAlaGlyGluTh 954

   1149 GCGCTCATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGG 1100
        |   |||              |||:::::   :::   |||   |||:
    954 rThrSerGlyCysSerProLysThrValThrThrThrValProThrThrT 971

   1099 CTGTCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTG 1050
        ::::::::::::   :::                :::         ::::::
    971 hrThrThrSerValThrThr............SerSerThrThrThrIle 983

   1049 CGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTC 1000
            :::           :::|||  :::::::::     ::::::||
    984 ThrThrThrValCysSerThrGlyThrAsnSerAlaGlyGluThrThrSe 1000

    999 TTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCA 950
        |            |||:::::||||||
   1000 rGlyCysSerProLysThrIleThr......................... 1008

    949 AATAATCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAA 900
              :::   :::||||||:::    :::|||    :::|||        |||
   1009 .....ThrThrValProCysSerThrSerProSerGluThrAlaSerGlu 1023

    900 .................................................. 900

   1024 SerThrThrThrSerProThrThrProValThrThrValValSerThrTh 1040

    899 .......ATCACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTT 857
                |||:::   |||:::||||||:::        :::
   1040 rValValThrThrGluTyrSerThrSerThrLysProGlyGlyGluIleT 1057

    856 GAGACACCTTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACT 807
           ||||||:::   |||::::::                   |||   |||
   1057 hrThrThrPheValThrLysAsnIle...ProThrThrTyrLeuThrThr 1072

    806 TGCAGGCCG...CCCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTGT 760
                 |||   |||:::::::||||||                   ||
   1073 IleAlaProThrProSerValThrThr.....................Va 1082

    759 TACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACA 710
        |||||||      |||    :::|||||||:::                  |
   1082 lThrAsnPheThrProThrThrIleThrThr..................T 1093

    709 CGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCGTGCCACTC 660
        |||||       :::   |||:::|||               :::
   1093 hrValCysSerThrGlyThrAsnSer...............AlaGlyGlu 1104

    659 AAGCCGGCAGGATGCGGGCCG............................. 639
              :::|||||||:::|||
   1105 ThrThrSerGlyCysSerProLysThrValThrThrThrValProCysSe 1121

    639 .................................................. 639

   1121 rThrGlyThrGlyGluTyrThrThrGluAlaThrThrLeuValThrThrA 1138

    639 .................................................. 639

   1138 laValThrThrThrValValThrThrGluSerSerThrGlyThrAsnSer 1154

    638 ......................................CCAAATTCATG 628
                                                |||::::::
   1155 AlaGlyLysThrThrThrGlyTyrThrThrLysSerValProThrThrTy 1171

    627 TGTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTT 578
           |||:::   ||||||        :::      |||     |||      |||:
   1171 rValThrThrLeuAlaProSerAlaProValThrProAlaThrAsnAlaV 1188

    577 TACACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCG 528
        ::    |||   |||        :::               ||||||
   1188 alProThrThrIleThrThrThrGluCysSerAlaAlaThrAsnAlaAla 1204

    527 CGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGC 478
                 :::|||::::::|||                |||:::||||||::
   1205 GlyGluThrThrSerValCysSerAlaLysThrIleValSerSerAlaSe 1221

    477 CAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCAT 428
        ::::|||                               :::::::::|||
   1221 rAlaGly...............................GluAsnThrAlaP 1228

    427 CTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAG 378
          :::|||          ::::::       :::     ::::::|||
   1228 roSerAlaThrThrProValThrThrAlaIleProThrThrVal...... 1242

    377 CCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAG 328
                           |||    |||    ::::::          ||
   1243 .................IleThrThrGluSerSerValGlyThrAsnSe 1253

    327 CGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCAACGG 278
        ||||:::   |||          :::   :::   :::|||:::|||:::|||
   1253 rAlaGlyGluThrThrThrGlyTyrThrThrLysSerIleProThrThrT 1270

    277 CAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCG 228
          |||||||||
   1270 yrIleThrThr....................................... 1273

    227 ATTTTGCCT........................GAAGCCGGCGCAGTAAA 202
        ::::::|||                          |||:::   |||:::
   1274 LeuIleProGlySerAsnGlyAlaLysAsnTyrGluThrValAlaThrAl 1290
```

```
  201 TACTACGCCC...........................GGATTCTTTTTGT 179
       |||:::|||                           :::        |
 1290 aThrAsnProIleSerIleLysThrThrSerGlnLeuAlaThrThrAlaS 1307

  178 CTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTTGATTTTC 129
       ||::::::|||:::   |||    :::|||   |||||||||||
 1307 erAlaSerSerValAlaProValValThr...SerProSerLeu...... 1320

  128 ATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTCGGT 79
                      |||     |||:::   ::::::|||::::::::
 1321 .........ThrGlyProLeuGlnSerAlaSerGlySerAlaValAlaTh 1334

   78 AATGGCCGGGCCGTCATAAATGACT...........TGCTCCGGTCTGC 41
       :   :::  ||||||      :::            :::::::::::
 1334 rTyrSerValProSerIleSerSerThrTyrGlnGlyAlaAlaAsnIleL 1351

   40 CCGCGATGGGCAGATTTAGACCTTTTTTGATTTTAATC 3
         ::::::|||:::|||      :::|||:::   :::
 1351 ysValLeuGlyAsnPheMetTrpLeuLeuLeuAlaLeu 1363
```

seq_name: SwissProt_40:MUC1_MOUSE

```
seq_documentation_block:
ID   MUC1_MOUSE     STANDARD;      PRT;   630 AA.
AC   Q02496;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mucin 1 precursor (Polymorphic epithelial mucin) (PEMT) (Episialin).
GN   MUC1 OR MUC-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91332029; PubMed=1714452;
RA   Spicer A.P., Parry G., Patton S., Gendler S.J.;
RT   "Molecular cloning and analysis of the mouse homologue of the tumor-
RT   associated mucin, MUC1, reveals conservation of potential O-
RT   glycosylation sites, transmembrane, and cytoplasmic domains and a
RT   loss of minisatellite-like polymorphism.";
RL   J. Biol. Chem. 266:15099-15109(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92068178; PubMed=1958179;
RA   Vos H.L., Devries Y., Hilkens J.;
RT   "The mouse episialin (Muc1) gene and its promoter: rapid evolution of
RT   the repetitive domain in the protein.";
RL   Biochem. Biophys. Res. Commun. 181:121-130(1991).
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC       CYTOSKELETON.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXCLUSIVELY LOCATED
CC       IN THE APICAL DOMAIN OF THE PLASMA MEMBRANE OF HIGHLY
CC       POLARIZED EPITHELIAL CELLS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF EPITHELIAL
CC       TISSUES. ABERRANTLY EXPRESSED IN EPITHELIAL CARCINOMAS.
CC   -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M84683; AAA39756.1; -.
DR   EMBL; U16175; AAA98538.1; -.
DR   EMBL; M65132; AAA39755.1; -.
DR   EMBL; M64928; AAA39755.1; JOINED.
DR   EMBL; M77226; AAA39754.1; -.
DR   PIR; A39344; A39344.
DR   MGD; MGI:97231; Muc1.
DR   InterPro; IPR000082; SEA.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW   Repeat.
FT   SIGNAL        1     20
FT   CHAIN        21    630       MUCIN 1.
FT   DOMAIN       21    535       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    536    556       POTENTIAL.
FT   DOMAIN      557    630       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       42    366       16 X 20 AA TANDEM APPROXIMATE REPEATS.
FT   REPEAT       42     61       1.
FT   REPEAT       62     81       2.
FT   REPEAT       82    101       3.
FT   REPEAT      102    122       4 (APPROXIMATE).
FT   REPEAT      123    143       5 (APPROXIMATE).
FT   REPEAT      144    164       6 (APPROXIMATE).
FT   REPEAT      165    184       7.
FT   REPEAT      185    204       8.
FT   REPEAT      205    225       9 (APPROXIMATE).
FT   REPEAT      226    246       10 (APPROXIMATE).
FT   REPEAT      247    256       11.
FT   REPEAT      257    286       12.
FT   REPEAT      287    306       13.
FT   REPEAT      307    326       14.
FT   REPEAT      327    346       15.
FT   REPEAT      347    366       16.
FT   DOMAIN      411    526       SEA.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    275    275       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    335    335       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    355    355       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    366    366       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    120    120       P -> L (IN REF. 2).
FT   CONFLICT    121    121       L -> S (IN REF. 2).
FT   CONFLICT    138    139       AT -> PA (IN REF. 2).
FT   CONFLICT    140    140       T -> TT (IN REF. 2).
FT   CONFLICT    423    423       F -> S (IN REF. 2).
FT   CONFLICT    506    506       S -> D (IN REF. 2).
FT   CONFLICT    602    602       Q -> S (IN REF. 2).
SQ   SEQUENCE   630 AA;  64622 MW;  FF57C1B31137C83B CRC64;
```

```
alignment_scores:
           Quality:  106.50            Length:    456
             Ratio:  0.520               Gaps:     21
 Percent Similarity:  44.956   Percent Identity: 23.246

alignment_block:
 US-09-303-518D-131/rev x MUC1_MOUSE   ..

 Align seg 1/1  to: MUC1_MOUSE  from: 1  to: 630

  1337 TCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTCGTA 1288
        |||    |||::::::|||:::|||
    38 SerSerSerLeuAlaSerThrThr.......................... 45

  1287 TTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATT 1238
                       |||    :::|||:::      |||:::|||:::   :
    46 .............ThrProValHisSer......SerAsnSerAspProA 56

  1237 CCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCG 1188
        ::   :::|||   :::    :::    |||    ||||||:::    :::|||
    56 laThrArgProProGlyAspSerThrSer...SerProValGlnSerSer 71
```

```
1187 CGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGT 1138
       :::|||  :::  |||      :::        |||  :::  ||
  72 ThrSerSerProAlaThrArgAlaProGluAspSerThrSerThrAlaVa 88

1137 GCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACT 1088
     |       ||||||          ||||||      ||||||  ||||||
  88 lLeuSerGlyThr......SerSerProAlaThrThrAlaProValAsnS 103

1087 TGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAG 1038
          |||         .                              :::
 103 erAlaSer.....................................Ser 106

1037 TATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCC 988
        ::::::            :::::::::|||   :::|||   |||:::
 107 ProValAlaHisGlyAspThrSerSerProAlaThrSerProLeuLysAs 123

 987 TTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCG 938
      |||:::   :::   :::
 123 pSerAsnSerSerProVal.............................. 129

 937 CGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTTG 888
                :::          :::::::::        |||
 130 ..........ValHisSerGlyThrSerSerAlaAlaThrThrAlaPro 142

 887 TCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAA 838
           ||||||:::|||||||:::|||            |||   :::|||
 143 ValAspSerThrSerSerProValValHisGlyGlyThrSerSerProAl 159

 837 AACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAA 788
      |||       :::     |||     |||     |||      ::::
 159 aThr......SerProProGlyAspSerThrSerSerProAspHisSerS 174

 787 CCACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCA 738
     ::|||   |||               |||:::|||   :::   |||
 174 erThrSerSer...........ProAlaThr...ArgAlaProGluAsp 185

 737 ATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGG 688
         ||||||                  |||::::::::::::   |||::
 186 SerThrSer.................ThrAlaValLeuSerGlyThrSe 196

 687 CTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGC 638
     :|||                            ||||||      :::|||
 196 rSer.........................ProAlaThrThrAlaProV 204

 637 CAAATTCATGTGTTTCGATATTGGCA....................... 612
        :::|||      |||    :::|||
 204 alAspSerThrSerSerProValAlaHisAspAspThrSerSerProAla 220

 611 ...GCATTTTCAGAC............................GGCAC 595
        ::::::|||:::                             |||||
 221 ThrSerLeuSerGluAspSerAlaSerSerProValAlaHisGlyGlyTh 237

 594 GTCTGCGCCTGCTGCTTTACACACATGGATTTACGTTCGGTCAGGCGGC 545
     ||||:::||||||:::                           |||     |||
 237 rSerSerProAlaThr....................SerProLeuArgA 247

 544 TCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATG 495
        :::|||:::    |||       :::|||::::::   :::    :::
 247 spSerThrSerSerProValHisSerSerAlaSerIleGlnAsnIleLys 263

 494 ACCGTAGGGTCGGCAGCCAGC...........GGATTGGTGTCCATCGC 457
      |||::::::::    ||||||                 :::   :::|||::::
 264 ThrThrSerAspLeuAlaSerThrProAspHisAsnGlyThrSerValTh 280

 456 ATTGACGAAGATGGCGAACGGCTCGGCA..................... 429
      :    |||:::     |||    |||||||||
 280 rThrThrSerSerAlaLeuGlySerAlaThrSerProAspHisSerGlyT 297

 428 ................TCTACGGCAGGGATTTTGCTGAACGGACGGGTG 396
                      |||:::     ::::::|||     :::       |||
 297 hrSerThrThrThrAsnSerSerGluSerValLeuAlaThrThrProVal 313

 395 CGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTC 346
        |||::::::                                  |||||
 314 TyrSerSerMet............................ProPheSe 320

 345 GCTGCTCAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCG...AACTCGA 299
     |         :::   :::|||:::|||:::               :::   |||:::
 320 rThrThrLysValThrSerGlySerAlaIleIleProAspHisAsnGlyS 337

 298 TTTCGTCGTTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTT 249
        |||    ||||||::::::::::::         :::
 337 erSerValLeuProThrSerSerValLeuGlySer.............. 348

 248 TCGCCACGGTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATAC 199
                   ::::::::::::    :::::::::|||:::    ||
 349 ............AlaThrSerLeuValTyrAsnThrSerAlaIleAlaTh 361

 198 TACGCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGG 149
     |||||||               ||||||:::|||     |||            :
 361 rThrPro...........ValSerAsnGlyThrGlnPro.........S 371

 148 CTTCACCTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCG 99
     ::    ||||||         ::::::      |||    |||    ||||||
 371 erValProSerGlnTyrProValSerProThrMetAlaThrThrSerSer 387

  98 CCAAGCAACGCGACTTCG 81
        |||:::   :::|||
 388 HisSerThrIleAlaSer 393
```

seq_name: SwissProt_40:ATPB_NICRU

```
seq_documentation_block:
ID   ATPB_NICRU     STANDARD;      PRT;   498 AA.
AC   P26530;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ATP synthase beta chain (EC 3.6.3.14).
GN   ATPB.
OS   Nicotiana rustica (Aztec tobacco).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4093;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92390714; PubMed=1387730;
RA   Avni A., Anderson J.D., Holland N., Rochaix J.-D., Gromet-Elhanan Z.,
RA   Edelman M.;
RT   "Tentoxin sensitivity of chloroplasts determined by codon 83 of beta
RT   subunit of proton-ATPase.";
RL   Science 257:1245-1247(1992).
CC   -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC       GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC       SUBUNIT.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
CC   -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X61318; CAA43611.1; -.
```

```
DR   PIR; S15724; PWNTBZ.
DR   HSSP; P00829; 1BMF.
DR   Mendel; 2237; NICru;atpB;1.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR004100; ATP-synt_ab_N.
DR   InterPro; IPR000793; ATPase_AB_C.
DR   InterPro; IPR000194; ATPase_alpha_beta.
DR   Pfam; PF00006; ATP-synt_ab; 1.
DR   Pfam; PF00306; ATP-synt_ab_C; 1.
DR   Pfam; PF02874; ATP-synt_ab_N; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW   ATP synthesis; Chloroplast; Thylakoid; Membrane; CF(1);
KW   Hydrolase; ATP-binding; Hydrogen ion transport.
FT   NP_BIND     172    179       ATP (BY SIMILARITY).
SQ   SEQUENCE   498 AA;  53470 MW;  2DE3CE1F95B76F2D CRC64;

alignment_scores:
           Quality:  105.00            Length:    492
             Ratio:  0.507               Gaps:     28
 Percent Similarity:  42.073   Percent Identity:  22.358

alignment_block:
 US-09-303-518D-131 x ATPB_NICRU   ..

 Align seg 1/1  to: ATPB_NICRU  from: 1  to: 498

     178 GACAAAAAGAATCCGGGC.......................GTAGTATT 203
          :::|||||||||||||||                        |||:::
      16 GluLysLysAsnProGlyArgValValGlnIleIleGlyProValLeuAs 32

     204 TACTGCGCCGGCTTCAGGCAAAATC...............GCCGCTATTC 238
          :::|||             ||||||:::          |||   :::
      32 pValAlaPheProProGlyLysMetProAsnIleTyrAsnAlaLeuValV 49

     239 ACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAA... 285
           :::|||:::       |||    |||    :::    :::|||   |||
      49 alGlnGlyArgAspSerValGlyGlnProIleAsnValAlaCysGluVal 65

     286 ............GGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGA 323
                     ||||||:::::::::                    ::
      66 GlnGlnLeuLeuGlyAsnAsnArgVal.....................Ar 75

     324 AGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTG......A 367
          :|||:::|||      :::::|||    :::    |||:::::      |
      75 gAlaValAlaMetSerAlaThrGluGlyLeuThrArgGlyMetGluValI 92

     368 TTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATC 417
          ||::::::|||                      |||:::|||   :::
      92 leAspThrGly.....................AlaProIleSer...Val 100

     418 CCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC................. 450
          |||:::       |||          ||||||
     101 ProValGlyGlyAlaThrLeuGlyArgIlePheAsnValLeuGlyAlaAl 117

     451 ...................GTCAATGCGATGGACACCAATCCGCTG.... 477
                             |||::::::      |||:::|||:::
     117 aValAspAsnLeuGlyProValAspThrSerThrThrSerProIleHisA 134

     478 ..GCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAA 525
            :::|||          :::    :::             :::   |||:::
     134 rgSerAlaProAlaPheIleGlnLeuAspThrLysLeuSerIlePheGlu 150

     526 CGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTG 575
            |||:::     |||::::::   |||
     151 ThrGlyIleLysValValAspLeuLeu....................... 159

     576 TAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAA 625
                                 :::|||    ::::::::::::::|||
     160 ...................AlaProTyrArgArgGlyGlyLysIleGlyL 170

     626 CACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATT 675
                                  |||||||||
     170 eu......PheGlyGly................................. 173

     676 CATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTA 725
                             :::|||:::::::||||||||||      ::::::
     174 ...............AlaGlyValGlyLysThrValLeuIleMetGluLe 185

     726 TCAAGACGTGATTGCT.................................A 742
             :::    ||||||                                :
     185 uIleAsnAsnIleAlaLysAlaHisGlyGlyValSerValPheGlyGlyV 202

     743 TCGGACGT.....................TTGTTCGTA............ 759
          :::|||:::                    |||::::::
     202 alGlyGluArgThrArgGluGlyAsnAspLeuTyrMetGluMetLysGlu 218

     760 ACAGGCCGTCTGAATACCGAGCGCGTG............GTTGCCTTGGG 797
          :::|||    :::|||    |||::::::            |||||||||
     219 SerGlyValIleAsnGluGluAsnIleAlaGluSerLysValAlaLeuVa 235

     798 CGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGA 847
                |||:::|||:::|||             |||      :::|||
     235 lTyrGlyGlnMetAsnGluProProGlyAlaArgMetArgValGly.... 250

     848 AGGTGTCTCAACTTACCGCCGGCGAA........................ 873
            ::::::    ||||||    :::|||
     251 ..LeuThrAlaLeuThrMetAlaGluTyrPheArgAspValAsnGluGln 266

     874 ......TTGGTTGACGCGGACAAC.........CGCGTGATTTCCGGTTC 908
                |||:::        ||||||          |||    :::||||||
     267 AspValLeuLeuPheIleAspAsnIlePheArgPheValGlnAlaGlySe 283

     909 GGTATTGAACGGTGCGATTGCACAA......GGCGCGCATGATTAT.... 948
          |    :::::::::   :::::::::          :::|||    |||
     283 rGluValSerAlaLeuLeuGlyArgMetProSerAlaValGlyTyrGlnP 300

     949 .................TTGGGACGCTACCACAATCAGATTTCCGTTATC 981
                            :::|||       :::::::::|||:::
     300 roThrLeuSerThrGluMetGlySerLeuGlnGluArgIleThrSerThr 316

     982 GAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGA 1031
          :::||||||              ::::::    :::       :::|||      ||
     317 LysGluGlySerIleThrSerIleGlnAlaValTyrValProAlaAspAs 333

    1032 CAAATACTCC...ATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACA 1078
          |      :::               ||||||:::::|||
     333 pLeuThrAspProAlaProAlaThrThrPheAlaHis............. 345

    1079 AACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTA 1128
            :::          ||||||:::::::::           |||:::::::::
     346 .....LeuAspAlaThrThrValLeuSer.........ArgGlyLeuAla 357

    1129 CCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTT 1178
               |||    |||         |||    |||||||||          |||
     358 AlaLysGlyIleTyrProAlaValAspProLeuAspSerThrSerThr.. 373

    1179 GCTTTTGCGCGATTTAATCGTCGGCGAT...................... 1206
           :::|||:::          |||||||||:::
     374 .MetLeuGlnProArgIleValGlyGluGluHisTyrGluThrAlaGlnA 390

    1207 ..................................ACCGACAGCGCGCAG 1221
                                                 |||
     390 rgValLysGlnThrLeuGlnArgTyrLysGluLeuGlnAspIleIleAla 406

    1222 GCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGC.. 1269
            ||||||        ||||||:::|||||||||         |||     :::
     407 IleLeuGlyLeuAspGluLeuSerGluGluAspArgLeuLeuValAlaAr 423

    1270 ....................................TTCGTCTGC.... 1278
                                              ||||||:::
     423 gAlaArgLysIleGluArgPheLeuSerGlnProPhePheValAlaGluV 440

    1279 ..............CCGGGCAAATAC 1290
                       ||||||||||||
```

```
    440 alPheThrGlySerProGlyLysTyr 448

seq_name: SwissProt_40:ATPB_TOBAC

seq_documentation_block:
ID   ATPB_TOBAC     STANDARD;      PRT;   498 AA.
AC   P00826;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ATP synthase beta chain (EC 3.6.3.14).
GN   ATPB.
OS   Nicotiana tabacum (Common tobacco).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. BRIGHT YELLOW 4;
RA   Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N.,
RA   Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA   Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.Y., Sugita M.,
RA   Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
RA   Tohdoh N., Shimada H., Sugiura M.;
RT   "The complete nucleotide sequence of the tobacco chloroplast genome:
RT   its gene organization and expression.";
RL   EMBO J. 5:2043-2049(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84059075; PubMed=6227526;
RA   Shinozaki K., Deno H., Kato A., Sugiura M.;
RT   "Overlap and cotranscription of the genes for the beta and epsilon
RT   subunits of tobacco chloroplast ATPase.";
RL   Gene 24:147-155(1983).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. XANTHI;
RX   MEDLINE=92390714; PubMed=1387730;
RA   Avni A., Anderson J.D., Holland N., Rochaix J.-D., Gromet-Elhanan Z.,
RA   Edelman M.;
RT   "Tentoxin sensitivity of chloroplasts determined by codon 83 of beta
RT   subunit of proton-ATPase.";
RL   Science 257:1245-1247(1992).
CC   -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC       GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC       SUBUNIT.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
CC   -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z00044; CAA77360.1; -.
DR   EMBL; K00507; AAA84676.1; -.
DR   EMBL; X61319; CAA43612.1; -.
DR   PIR; A01027; PWNTB.
DR   HSSP; P00829; 1BMF.
DR   Mendel; 4300; NICta;atpB;1.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR004100; ATP-synt_ab_N.
DR   InterPro; IPR000793; ATPase_AB_C.
DR   InterPro; IPR000194; ATPase_alpha_beta.
DR   Pfam; PF00006; ATP-synt_ab; 1.
DR   Pfam; PF00306; ATP-synt_ab_C; 1.
DR   Pfam; PF02874; ATP-synt_ab_N; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW   ATP synthesis; Chloroplast; Thylakoid; Membrane; CF(1);
KW   Hydrolase; ATP-binding; Hydrogen ion transport.
FT   NP_BIND     172    179       ATP (BY SIMILARITY).
SQ   SEQUENCE   498 AA;  53554 MW;  87A17E1185EB2CF3 CRC64;

alignment_scores:
          Quality:  105.00           Length:    492
            Ratio:  0.507              Gaps:     28
 Percent Similarity:  42.073   Percent Identity:  22.358

alignment_block:
 US-09-303-518D-131 x ATPB_TOBAC   ..

 Align seg 1/1  to: ATPB_TOBAC  from: 1  to: 498

    178 GACAAAAAGAATCCGGGC.......................GTAGTATT 203
        ::::|||||||||||||||                      |||:::
     16 GluLysLysAsnProGlyArgValValGlnIleIleGlyProValLeuAs 32

    204 TACTGCGCCGGCTTCAGGCAAAATC..............GCCGCTATTC 238
         :::|||         ||||||:::              |||    :::
     32 pValAlaPheProProGlyLysMetProAsnIleTyrAsnAlaLeuValV 49

    239 ACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAA... 285
          :::|||:::       |||   |||   :::   :::|||   |||
     49 alGlnGlyArgAspSerValGlyGlnProIleAsnValAlaCysGluVal 65

    286 ...........GGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGA 323
                   ||||||:::::::::                    ::
     66 GlnGlnLeuLeuGlyAsnAsnArgVal...................Ar 75

    324 AGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTG......A 367
        :|||:::|||       :::::|||   :::   |||::::::      |
     75 gAlaIleAlaMetSerAlaThrGluGlyLeuThrArgGlyMetGluValI 92

    368 TTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATC 417
        |||:::::|||                        |||:::|||   :::
     92 leAspThrGly.....................AlaProIleSer...Val 100

    418 CCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC................. 450
        |||:::        |||           ||||||
    101 ProValGlyGlyAlaThrLeuGlyArgIlePheAsnValLeuGlyGluPr 117

    451 ..................GTCAATGCGATGGACACCAATCCGCTG.... 477
                          |||:::::        |||:::|||:::
    117 oValAspAsnLeuGlyProValAspThrSerThrThrSerProIleHisA 134

    478 ..GCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAA 525
         .:::|||        :::   :::            :::   |||:::
    134 rgSerAlaProAlaPheIleGlnLeuAspThrLysLeuSerIlePheGlu 150

    526 CGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTG 575
           |||:::   |||::::::   |||
    151 ThrGlyIleGluValValAspLeuLeu....................... 159

    576 TAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAA 625
                              :::|||   :::::::::::::::|||
    160 ..................AlaProTyrArgArgGlyGlyLysIleGlyL 170

    626 CACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATT 675
                  |||||||||
    170 eu......PheGlyGly................................ 173

    676 CATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTA 725
                    :::|||::::::|||||||||       ::::::
```

```
 174 ..............AlaGlyValGlyLysThrValLeuIleMetGluLe 185

 726 TCAAGACGTGATTGCT...............................A 742
         :::   ||||||                                :
 185 uIleAsnAsnIleAlaLysAlaHisGlyGlyValSerValPheGlyGlyV 202

 743 TCGGACGT....................TTGTTCGTA............ 759
     ::|||:::                    |||::::::
 202 alGlyGluArgThrArgGluGlyAsnAspLeuTyrMetGluMetLysGlu 218

 760 ACAGGCCGTCTGAATACCGAGCGCGTG...........GTTGCCTTGGG 797
      :::|||   :::|||  |||::::::          |||||||||
 219 SerGlyValIleAsnGluGluAsnIleAlaGluSerLysValAlaLeuVa 235

 798 CGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTGGGTGCGA 847
           |||:::|||:::|||          |||      :::|||
 235 lTyrGlyGlnMetAsnGluProProGlyAlaArgMetArgValGly.... 250

 848 AGGTGTCTCAACTTACCGCCGGCGAA....................... 873
       ::::::   ||||||   ::::|||
 251 ..LeuThrAlaLeuThrMetAlaGluTyrPheArgAspValAsnGluGln 266

 874 ......TTGGTTGACGCGGACAAC.........CGCGTGATTTCCGGTTC 908
          |||:::     ||||||           |||   :::||||||
 267 AspValLeuLeuPheIleAspAsnIlePheArgPheValGlnAlaGlySe 283

 909 GGTATTGAACGGTGCGATTGCACAA......GGCGCGCATGATTAT.... 948
     |    :::::::::   :::::::::       :::|||     |||
 283 rGluValSerAlaLeuLeuGlyArgMetProSerAlaValGlyTyrGlnP 300

 949 ................TTGGGACGCTACCACAATCAGATTTCCGTTATC 981
                     :::|||      :::::::::|||:::
 300 roThrLeuSerThrGluMetGlySerLeuGlnGluArgIleThrSerThr 316

 982 GAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGA 1031
     :::||||||           ::::::   :::     :::|||     ||
 317 LysGluGlySerIleThrSerIleGlnAlaValTyrValProAlaAspAs 333

1032 CAAATACTCC...ATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACA 1078
     |       :::            ||||||:::::|||
 333 pLeuThrAspProAlaProAlaThrThrPheAlaHis............ 345

1079 AACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTA 1128
           :::      ||||||:::::::::         |||:::::::::
 346 .....LeuAspAlaThrThrValLeuSer.........ArgGlyLeuAla 357

1129 CCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTT 1178
         |||   |||      |||   |||||||||        |||
 358 AlaLysGlyIleTyrProAlaValAspProLeuAspSerThrSerThr.. 373

1179 GCTTTTGCGCGATTTAATCGTCGGCGAT..................... 1206
      :::|||:::      |||||||||:::
 374 .MetLeuGlnProArgIleValGlyGluGluHisTyrGluThrAlaGlnA 390

1207 .................................ACCGACAGCGCGCAG 1221
                                            |||
 390 rgValLysGlnThrLeuGlnArgTyrLysGluLeuGlnAspIleIleAla 406

1222 GCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGC.. 1269
        ||||||      ||||||:::|||||||||     |||    :::
 407 IleLeuGlyLeuAspGluLeuSerGluGluAspArgLeuLeuValAlaAr 423

1270 ..................................TTCGTCTGC.... 1278
                                        ||||||:::
 423 gAlaArgLysIleGluArgPheLeuSerGlnProPhePheValAlaGluV 440

1279 .............CCGGGCAAATAC 1290
                  ||||||||||||
 440 alPheThrGlySerProGlyLysTyr 448
```

seq_name: SwissProt_40:PPSA_PYRHO

seq_documentation_block:

```
ID   PPSA_PYRHO     STANDARD;      PRT;   821 AA.
AC   O57830;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate,water
DE   dikinase) (PEP synthase).
GN   PPSA OR PH0092.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC       phosphoenolpyruvate + phosphate.
CC   -!- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC       LACTATE ARE USED AS A CARBON SOURCE.
CC   -!- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AP000001; BAA29161.1; -.
DR   HSSP; P22983; 1DIK.
DR   InterPro; IPR000121; PEP_utilizers.
DR   InterPro; IPR002192; PPDK_N_term.
DR   Pfam; PF00391; PEP-utilizers; 1.
DR   Pfam; PF02896; PEP-utilizers_C; 1.
DR   Pfam; PF01326; PPDK_N_term; 1.
DR   ProDom; PD000940; PEP_utilizers; 1.
DR   PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR   PROSITE; PS00742; PEP_ENZYMES_2; 1.
KW   Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.
FT   MOD_RES     444    444       PHOSPHORYLATION (BY SIMILARITY).
FT   DOMAIN      812    818       POLY-GLU.
SQ   SEQUENCE   821 AA;  90812 MW;  A93816D865F8A0BF CRC64;
```

```
alignment_scores:
           Quality:  105.00          Length:    477
             Ratio:  0.438             Gaps:     21
  Percent Similarity:  50.314   Percent Identity: 21.174

alignment_block:
 US-09-303-518D-131 x PPSA_PYRHO   ..

 Align seg 1/1  to: PPSA_PYRHO  from: 1  to: 821

   31 CCCATCGCGGGCAGACCGGAGCAAGTCATTTATGAC............. 66
      |||:::::::::::   ::::::::::::   :::
  238 ProValThrAsnAsnArgAsnGluIleMetIleAsnAlaSerTrpGlyLe 254

   67 .GGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCA 115
        |||    |||:::::::::   |||:::      :::||||||:::
  254 uGlyGluAlaValValSerGlyAlaValThrProAspGluTyrIleValG 271

  116 TGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAA 165
```

```
           :::   :::   |||||||||||   :::::::::   |||   :::
271 luLysGlyThrTrpLysIleLysGlu...LysValIleAlaLysLysGlu 286

166 GTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGC 215
     |||:::                                 |||:::      |||
287 ValMet.........................ValIleArgAsnProGl 294

216 TTCAGGCAAA...............ATCGCCGCTATTCACCGTGGC.... 246
      :::||||||                  :::|||   :::   |||
294 uThrGlyLysGlyThrValGlnValLysValAlaGluTyrLeuGlyProG 311

247 ........GAAAAGCGCGTACTT.........CAGTCAGTCGTGATTGCC 279
                 ||||||:::||||||          |||   :::  :::|||
311 luTrpValGluLysGlnValLeuThrAspGluGlnIleIleGluValAla 327

280 GTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCT 329
             |||:::          |||         :::      :::
328 LysMetGlyGlnLysIleGluGluHisTyrGlyTrpProGlnAspIleGl 344

330 GGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCT 379
               :::::::::      :::      ::::::||||||
344 uTrpAlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerArgP 361

380 TATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGAT 429
         |||:::|||:::                :::::::::       |||
361 roIleThrThrLeuLysGluThrThrThrGluGluValGluGluVal... 376

430 GCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG...CT 476
        |||            ::::::::::::::::::      ::::::|||    ::
377 ...GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyIl 392

477 GGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAA......GACT 520
     ::::|||         :::|||:::|||    :::|||:::|||
392 eGlyAlaGlyArgValValValIlePheAspAlaSerGluIleAspLysV 409

521 TCAAACGCGGC...CTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATC 567
      |||:::|||    :::||||||    :::         :::           :::
409 alLysGluGlyAspValLeuValThrThrMetThrAsnProAspMetVal 425

568 CATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCT..... 612
       :::    |||    |||:::|||    :::    :::|||:::::::
426 ProAlaMetLysArgAlaSerAlaIleIleThrAspGluGlyGlyArgTh 442

613 .......GCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCG 655
             |||    |||    :::::::|||:::|||    |||         :::|
442 rSerHisAlaAlaIleValSerArgGluLeuGlyIleProAlaValValG 459

656 GCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAA 705
     ||    :::                                       |||:::|||
459 lyThrLys............................GluAlaThrLys 465

706 ACCGTGTGGACCATCAATTATCAAGACGTG.................... 735
       :::    |||    :::|||       |||
466 LysLeuLysThrGlyAspTyrValThrValAspGlyThrArgGlyLeuVa 482

736 ....ATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGC 781
            :::|||    :::    :::|||        :::    :::    |||
482 lTyrLysGlyIleValLysSerLeuValGluLysLysLysLysGluGluA 499

782 GCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGT 831
      ::::::|||    |||:::        :::        |||    :::
499 laAlaAlaAlaProGlyAlaAlaValAlaAlaAlaProLeuValThrGly 515

832 ACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGA 881
     |||::::::    ::::::||||||:::          ::::::|||
516 ThrLeuValLysValAsnValSerMetProGluValAlaGlu........ 529

882 CGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCAC 931
                |||:::    :::    ::::::    :::    ::::::    :
530 ..........ArgAlaAlaAlaThrGlyAlaAspGlyValGlyLeuLeuA 543

932 AAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATC 981
     ::::    |||        |||:::        ::::::    ::::::    |||
543 rgAlaGluHisMetIleLeuSerIleGlyGlnHisProValLysPheIle 559

982 GAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGA 1031
     :::||||||::::::    ||||||            :::|||         ::
560 LysGluGlyLysGluAspGluLeuValGluArgLeuAlaGluGlyIleGl 576

1032 CAAATACTCCATCACGCGCACCACTCTCGGCCATTT.............. 1067
      :|||    :::       |||    :::    ||||||
576 uLysValAlaAla..AlaPheTyrProArgProValTrpTyrArgThrLe 592

1068 .......CCTAAAAAACAAACTCTTCAAGTT.........CACGACAGCC 1101
                |||    ::::::       ::::::          :::|||:::
592 uAspAlaProThrAsnGluPheArgGluMetProGlyGlyGluAspGluP 609

1102 GTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGT 1151
       ::::::|||:::|||         |||                        |||
609 roGluGluArgAsnProMetLeuGlyTrpArgGly.........IleArg 622

1152 AATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCG 1201
      ::::::::    :::|||       |||       |||:::|||:::
623 ArgGlyLeuAspGlnProGluLeuLeuArg...AlaGluPheLysAlaIl 638

1202 GCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAA 1251
       :::          :::    |||::::::    :::|||
638 eLysLysValValGluLysGlyTyrAsnAsnIleGly............ 650

1252 GACCTCGCTTTGTGCAGCTTCGTCTG.........CCCGGGCAAATACGA 1292
                    |||:::|||    |||          |||    |||||||||
651 ..........ValMetLeuProLeuValSerHisProGluGlnIleArg 663

1293 ATACGGCCCGCTGTTGCGCAAAGTGCTGGA 1322
                :::|||:::::::::|||
664 LysAlaLysGluIleAlaArgGluValGly 673
```

seq_name: SwissProt_40:SMB2_MESAU

```
seq_documentation_block:
ID   SMB2_MESAU     STANDARD;      PRT;   989 AA.
AC   Q60560;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA-binding protein SMUBP-2 (Immunoglobulin MU binding protein 2)
DE   (SMUBP-2) (Insulin II gene enhancer-binding protein) (RIPE3B-binding
DE   complex 3B2 P110 subunit) (RIP-1).
GN   IGHMBP2 OR SMUBP2 OR RIP1.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Insulinoma;
RX   MEDLINE=95394901; PubMed=7665561;
RA   Shieh S.-Y., Stellrecht C.M.M., Tsai M.-J.;
RT   "Molecular characterization of the rat insulin enhancer-binding
RT   complex 3b2. Cloning of a binding factor with putative helicase
RT   motifs.";
RL   J. Biol. Chem. 270:21503-21508(1995).
CC   -!- FUNCTION: BINDS TO THE INSULIN II GENE RIPE3B ENHANCER
CC       REGION.
CC   -!- SUBUNIT: COMPOSED OF AT LEAST THREE POLYPEPTIDES: P58, P62, AND
CC       P110.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: HIGH EXPRESSION IN BRAIN AND TESTIS, MODERATE
CC       IN HEART, SPLEEN, AND KIDNEY, AND LOW IN OTHER TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE DNA2/NAM7 HELICASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L15625; AAB00104.1; -.
DR   InterPro; IPR001374; R3H.
DR   InterPro; IPR000058; ZnF_AN1.
DR   Pfam; PF01424; R3H; 1.
DR   SMART; SM00393; R3H; 1.
DR   SMART; SM00154; ZnF_AN1; 1.
KW   Helicase; ATP-binding; DNA-binding; Nuclear protein;
KW   Transcription regulation; Activator.
FT   NP_BIND     213    220       ATP (POTENTIAL).
FT   DOMAIN      250    425       LEU-RICH.
FT   DOMAIN      370    373       POLY-VAL.
FT   DNA_BIND    637    783       SS DNA-BINDING (BY SIMILARITY).
FT   DOMAIN      793    859       GLN/PRO-RICH.
FT   DOMAIN      860    864       POLY-LYS.
FT   DOMAIN      860    864       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ   SEQUENCE   989 AA;  108439 MW;  9489671E46DAD04E CRC64;

alignment_scores:
           Quality:  105.00            Length:    473
             Ratio:  0.465               Gaps:     27
 Percent Similarity:  47.780   Percent Identity: 21.987

alignment_block:
 US-09-303-518D-131 x SMB2_MESAU   ..

 Align seg 1/1   to: SMB2_MESAU   from: 1   to: 989

    166 GTGCTGTTTGAAGACAAAAAG...AATCCGGGCGTAGTATTTACTGCGCC 212
         |||   ||||||   :::|||   :::|||   ||||||:::   :::
     65 ValThrPheGluProArgLysLeuGlyProValValValLeuProSerAs 81

    213 GGCT......TCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAG.....252
          :::        ||||||   |||::::::::::   :::::::::
     81 nSerPheThrSerGlyAspIleValGlyLeuTyrAspAlaAsnGluSerS 98

    253 .........................CGCGTACTTCAG...TCAGTCGTG 273
                                 |||:::   |||   |||||||:::
     98 erGlnLeuAlaThrGlyValLeuThrArgIleThrGlnLysSerValThr 114

    274 ATTGCCGTT...GAAGGCAACGAC...GAAATCGAGTTCGAACGCTACGT 317
         :::|||       |||::::::|||   :::::::::::::::|||
    115 ValAlaPheAspGluSerHisAspPheGlnLeuAsnLeuAspArgGluAs 131

    318 ACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGA 367
                 |||   ||||||:::::::   :::   ::::::|||
    131 nThrTyrArgLeuLeuLysLeuAlaAsnAspValThrTyrLysArgLeu. 147

    368 TTCAATCAGGCTTATGGACTGCGCTTCGC......ACCCGTCCGTTCAGC 411
           :::::::::|||   |||       :::      :::   |||   |||
    148 ..LysLysAlaLeuMetThrLeuLysLysTyrHisSerGlyProAlaSer 163

    412 AAAATC................................CCTGCCGTAGA 428
        ::::::                                |||::::::::
    164 SerLeuIleAspValLeuLeuGlyGlySerSerProSerProThrThrGl 180

    429 TGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAAT....... 471
         :       ||||||:::::::::   :::|||:::|||     :::
    180 uIleProProPheThrPheTyrAsnThrAlaLeuAspProSerGlnLysG 197

    471 .................................................. 471

    197 luAlaValSerPheAlaLeuAlaGlnLysGluValAlaIleIleHisGly 213

    472 ...CCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGA 518
            |||    ::::::       |||::::::::::::         :::
    214 ProProGlyThrGlyLysThrThrThrValValGluIleIleLeuGlnAl 230

    519 CTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCC 568
         |||:::||||||   :::|||   ::::::   |||:::         :::
    230 aValLysGlnGlyLeuLysIleLeu.CysCysAlaProSer......Asn 244

    569 ATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAAT 618
        ::::::|||   :::   ::::::         ||||||   :::  |||:::
    245 ValAlaValAspAsnLeuValGlu......ArgLeuAlaLeu.CysLysL 259

    619 ATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCAC 668
          |||             :::   :::||||||           :::   :::
    259 ysArgIleLeuArgLeuGlyHisProAlaArgLeuLeuGluSerAlaGln 275

    669 GCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCA 718
           ||||||:::       :::      |||:::   |||
    276 GlnHisSerLeuAspAlaValLeuAlaArgSerAspAsnAlaGlnIleVa 292

    719 TCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGT 768
                  |||:::|||       ::::::|||   :::|||:::    :
    292 lAlaAspIleArgLysAspIleAspGlnValPheGlyLysAsnLysLysT 309

    769 CTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACC 818
        :::::    :::|||        :::                  |||:::
    309 hrGln.AspLysArgGluLysSerAsnPheArg.........AsnGluIl 322

    819 GCGCCTCTTGCGTACCGTTTTG..............GGTGCGAAGGTGT 853
          :::|||||||||        |||                :::|||    |||:
    322 eLysLeuLeuArgLysGluLeuLysGluArgGluGluAlaAlaIleValG 339

    854 CTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCC 903
        :::::||||||||||:::::::::|||    |||    |||
    339 lnSerLeuThrAlaAlaAspValValLeuAlaThrAsn............ 351

    904 GGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGG 953
                    :::||||||   ::::::|||              |||
    352 ............ThrGlyAlaSerSerAspGlyProLeuLysLeuLeuPr 364

    954 ACGCTACCACAATCAGATTTCCGTTATCGAAGAA...GGCCGCAGCAAAG 1000
          :::    |||   ::::::    |||::::::|||    :::::::::    |
    364 oGluAsnHisPheAspValValValValAspGluCysAlaGlnAlaLeuG 381

   1001 AGCTGTTCGGCTGGGTT............GCGCCGCAG............ 1026
        ||            |||:::                 ||||||:::
    381 luAlaSerCysTrpIleProLeuLeuLysAlaProLysCysIleLeuAla 397

   1027 CCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAA 1076
           |||         ::::::           ||||||::::::|||
    398 GlyAspHisArgGlnLeuProProThrThrIleSerHis........... 410

   1077 CAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGG 1126
                           :::|||:::    |||    :::|||:::::::
    411 ................LysAlaAlaLeuAlaGlyLeuSerArgSerLeuM 422

   1127 TACCGATCGGCACTTATGAGCGCGTAATG..................... 1155
        ::                 ||||||::::::
    422 et..............GluArgLeuValGluLysHisGlyAlaGlyAla 433

   1155 .................................................. 1155

    434 ValArgMetLeuThrValGlnTyrArgMetHisGlnAlaIleThrArgTr 450

   1156 ........................CCGTTGGACATCCTGCCTACCTTG. 1179
                                        |||          |||::::::
    450 pAlaSerGluAlaMetTyrHisGlyGlnLeuThrAlaHisProSerValA 467

   1180 ........CTTTTGCGCGATTTAATC...GTCGGCGATACCGACAGCGCG 1218
                    ||||||:::||||||        |||:::||||||:::::::::
    467 laGlyHisLeuLeuLysAspLeuProGlyValAlaAspThrGluGluThr 483

   1219 CAG.........................GCTTTGGGTTGCTTGGAATT 1241
        :::                                :::    |||    ||||||||
    484 SerValProLeuLeuLeuIleAspThrAlaGlyCysGlyLeuLeuGluLe 500

   1242 GGACGAAGAAGAC 1254
        |||||||||||||
```

```
    500 uAspGluGluAsp 504

seq_name: SwissProt_40:ILVE_BACSU

seq_documentation_block:
ID   ILVE_BACSU     STANDARD;      PRT;   362 AA.
AC   P39576;
DT   01-FEB-1995 (Rel. 31, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative branched-chain amino acid aminotransferase (EC 2.6.1.42)
DE   (BCAT) (Vegetative protein 85) (VEG85).
GN   YWAA OR IPA-0R.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Glaser P., Lubochinsky B., Danchin A.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-57 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95020537; PubMed=7934828;
RA   Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA   Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA   Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA   Rapoport G., Danchin A.;
RT   "Bacillus subtilis genome project: cloning and sequencing of the 97
RT   kb region from 325 degrees to 333 degrees.";
RL   Mol. Microbiol. 10:371-384(1993).
RN   [3]
RP   SEQUENCE OF 1-9.
RC   STRAIN=IS58;
RX   MEDLINE=97443988; PubMed=9298659;
RA   Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA   Hecker M.;
RT   "First steps from a two-dimensional protein index towards a response-
RT   regulation map for Bacillus subtilis.";
RL   Electrophoresis 18:1451-1463(1997).
CC   -!- CATALYTIC ACTIVITY: L-LEUCINE + 2-OXOGLUTARATE = 4-METHYL-2-
CC       OXOPENTANOATE + L-GLUTAMATE (ALSO ACTS ON L-ISOLEUCINE AND
CC       L-VALINE).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: VALINE AND ISOLEUCINE BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z49992; CAA90289.1; -.
DR   EMBL; X73124; CAA51556.1; -.
DR   EMBL; Z99123; CAB15881.1; -.
DR   PIR; S39655; S39655.
DR   HSSP; P00510; 1A3G.
DR   SubtiList; BG10546; ywaA.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; Complete proteome.
FT   INIT_MET      0      0
FT   BINDING     196    196       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   362 AA;  40191 MW;  FFBA681A5FDC7057 CRC64;
```

```
alignment_scores:
           Quality:  104.50              Length:    328
             Ratio:  0.661                 Gaps:     15
 Percent Similarity:  48.171   Percent Identity:  22.561

alignment_block:
 US-09-303-518D-131 x ILVE_BACSU   ..

 Align seg 1/1   to: ILVE_BACSU   from: 1   to: 362

    160 GGCCAAGTGCTGTTTGAAGACAAAAAG..............AATCCGGG 194
        ||||||:::::||||||         |||               :::
     66 GlyGlnThrValPheGluGlyLeuLysAlaTyrValSerGluAspAspHi 82

    195 CGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTG 244
        |||::::::       |||                        ::::::
     82 sValLeuLeuPheArgPro.......................GluLysA 91

    245 GCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGAC 294
        ::   :::|||:::   ||||||                    ||||||
     91 snMetGluArgLeuAsnGlnSer.....................AsnAsp 100

    295 GAAATCGAGTTC............GAACGCTACGTACCTGAAGCGCTGGC 332
        ::::::   :::             |||:::   |||   |||::::|||
    101 ArgLeuCysIleProGlnIleAspGluGluGlnValLeuGluGlyLeuLy 117

    333 AAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTAT 382
          :::|||   :::        :::   ::::::   |||   ::::::
    117 sGlnLeuValAla......IleAspLysAspTrpIleProAsnAlaGluG 132

    383 GGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCC 432
          |||:::|||      |||||||||              :::   :::
    132 lyThrSerLeuTyrIleArgProPhe............IleIleAlaThr 144

    433 GAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGC 482
        |||||||||    :::    |||   |||      |||       |||
    145 GluProPhe...LeuGlyValAlaAlaSerHisThrTyrLysLeu..... 158

    483 CGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCC 532
                 :::|||:::::   :::::::::   :::|||:::|||:
    159 .......LeuIleIleLeuSerProValGlySerTyrTyrLysGluGlyI 173

    533 TGTTGGTATTGAGCCGCCTGACCGAACGTAAAATC............... 567
        ::       ::::::        :::|||    ::::::
    173 leLysProValLysIleAlaValGluSerGluPheValArgAlaValLys 189

    568 .........CATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAA 608
                    :::|||:::||||||          :::    |||    ::
    190 GlyGlyThrGlyAsnAlaLysThrAlaGlyAsnTyrAlaSerSerLeuLy 206

    609 TGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCT 658
        :|||    ::::::    :::::|||   |||         :::       |
    206 sAlaGlnGlnVal...AlaGluGluLysGlyPheSerGlnValLeuTrpL 222

    659 TGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACC 708
        ||:::|||    :::       :::||||||   |||||||:::    :::
    222 euAspGlyIleGluLysLysTyrIleGluGluValGlySerMetAsnIle 238

    709 GTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGT 758
           :::   ||||||   :::                           :::||
    239 PhePheLysIleAsnGlyGlu........................IleVa 247

    759 AACAGGCCGTCTGAAT.................................. 774
        ||||       ||||||
    247 lThrProMetLeuAsnGlySerIleLeuGluGlyIleThrArgAsnSerV 264

    775 ..................................ACCGAGCGCGTGGTT 789
                                           :::||||||    :::
    264 alIleAlaLeuLeuLysHisTrpGlyLeuGlnValSerGluArgLysIle 280

    790 GCCTTGGGCGGC...CTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGT 836
```

```
         |||:::          :::|||:::::|||       :::|||:::   ::
281 AlaIleAspGluValIleGlnAlaHisLysAspGlyIleLeuGluGluAl 297

837 TTTGGGTGCGAAGGTGTCTCAACTTACCGCC......GGCGAATTGGTTG 880
    ::::|||:::  ::::::  :::   :::       |||||||||:::
297 aPheGlyThrGlyThrAlaAlaValIleSerProValGlyGluLeuIleT 314

881 ACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCA 930
             |||:::   :::      ::::::    :::|||  ||||||
314 rpGlnAspGluThrLeuSerIleAsnAsnGlyGluThrGlyGluIleAla 330

931 CAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTAT 980
    :::      :::|||                          |||:::  ||
331 LysLysLeuTyrAsp....................ThrIleThrGlyIl 340

981 CGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGG 1014
    |::::::|||          :::   |||||||||
340 eGlnLysGlyAlaValAlaAspGluPheGlyTrp 351
```

seq_name: SwissProt_40:MUC1_MESAU

```
seq_documentation_block:
ID   MUC1_MESAU     STANDARD;      PRT;   676 AA.
AC   Q60528;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mucin 1 precursor.
GN   MUC1.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Tracheal epithelium;
RX   MEDLINE=96326118; PubMed=8703480;
RA   Park H., Hyun S.W., Kim K.C.;
RT   "Expression of MUC1 mucin gene by hamster tracheal surface epithelial
RT   cells in primary culture.";
RL   Am. J. Respir. Cell Mol. Biol. 15:237-244(1996).
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC       CYTOSKELETON (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U36918; AAB53965.1; -.
DR   InterPro; IPR000082; SEA.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW   Repeat.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    676       MUCIN 1.
FT   DOMAIN       26    582       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    583    603       POTENTIAL.
FT   DOMAIN      604    676       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      458    573       SEA.
FT   CARBOHYD    291    291       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    323    323       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    380    380       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    479    479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    496    496       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   676 AA;  67616 MW;  95F479B6EC5C3884 CRC64;

alignment_scores:
          Quality:  104.50             Length:    431
            Ratio:  0.515                Gaps:     22
 Percent Similarity:  47.100   Percent Identity: 23.202

alignment_block:
 US-09-303-518D-131/rev x MUC1_MESAU   ..

 Align seg 1/1  to: MUC1_MESAU  from: 1  to: 676

1283 CCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAA 1234
     |||:::::::::   |||    :::        |||||||||||
  21 ProAsnSerValAlaLeuSerGlnAspThrSerSerSerSer........ 34

1233 GCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCGCGCA 1184
                     :::|||::::::::|||:::    ::::::
  35 ................ThrLeuAsnThrThrProValHisSerGly.... 44

1183 AAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCG 1134
       |||::::::           |||:::::::::        :::    :::|||
  45 ..SerSerAlaProAlaThrSerSerAlaValAspSerAlaThrThrPro 60

1133 ATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTGAA 1084
            :::    :::    :::||||||   :::::::::||||||
  61 GlyHisSerGlySerSerAlaProProThrSerSerAlaValAsnSerAl 77

1083 GAGTTTGTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATT 1034
      :::
  77 aThrThr.........................................P 80

1033 TGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCT 984
      :::        |||::::::   |||    :::|||           |||
  80 roGlyHisSerGlySerSerAlaProProThrSerSerAlaValAsnSer 96

 983 TCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCC 934
     :::    |||    :::                   :::    |||   |||||
  97 AlaThrThrProVal..............HisSerGlySerSerAlaPr 108

 933 TTGTGCAATCGCACCGTTCAAT......ACCGAACCGGAAATCACGCGGT 890
     |   :::   :::        |||       |||    |||        :::
 108 oValThrSerSerAlaValAsnSerAlaThrThrProValHisSerGlyS 125

 889 TGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCC 840
       ||||||       :::|||    |||||||:::
 125 erSerAlaProProThrSerSerAlaValAsn.................. 135

 839 AAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGC 790
                                              :::||
 136 ..........................................SerAl 137

 789 AACCACG......CGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTC 746
     |||||||      :::|||              |||||||||::::::
 137 aThrThrProValHisSerGlySerSerAlaProValThrSerSer.... 152

 745 CGATAGCAATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCG 696
           |||::::::|||               ||||||:::         |||
 153 .....AlaValAsnSerAlaThrThrProValHisSerGlySerSerAla 167

 695 CCG......ACTGGCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGC 652
     |||      ::::::::::::               :::|||:::      ::
 168 ProValThrSerSerAlaValAspSerAlaThrThrProValHisSerGl 184

 651 AGGATGCGGGCCGCCAAATTCATGTGTT......TCGATATTGGCAGCAT 608
     ::::   :::||||||:::|||   :::      |||       :::
```

```
184 ySerSerAlaProProThrSerSerAlaValAsnSerAlaThrThrProV 201

607 TTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGT 558
           :::|||:::||||||||
201 alHisSerGlySerSerAlaPro.......................... 208

557 TCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGC 508
                                          :::   |||||||||::
209 ..................................ValThrSerSerAlaVa 214

507 TTCTTTGATGATGACC.........GTAGGGTCGGCAGCC.........A 476
    ::::         |||         ||||||:::|||          |
214 lAsnSerAlaThrThrProValHisSerGlySerSerAlaProValThrS 231

475 GCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCT 426
    ||:::   |||:::   |||   |||   :::   :::||||||::::::
231 erSerAlaValAsnSerAlaThrThrProValHisSerGlySerSerAla 247

425 ......ACGGCAGGGATTTTGCTGAACGGACGG......GTGCGAAGCGC 388
          |||::::::::::::   ::::::         |||:::|||::
248 ProProThrSerSerValValAsnSerAlaThrThrProValHisSerGl 264

387 AGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCA 338
    :           |||                   |||   |||   :::|
264 ySerSerAlaProPro.................ThrSerSerAlaValA 275

337 ATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTG 288
    ||:::|||::::::                   :::|||   ||||||
275 snLeuAlaThrThrPro............ValHisSerGlySerSerThr 287

287 CCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTG 238
    |||:::|||        |||||||||    :::|||      :::|||
288 ProAlaThrAsnSerThrThrAspSerAlaThr......ThrPro..... 300

237 AATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGAT 188
                 :::|||    :::::::::::::::||||||   ::::
301 .............ValProProGlySerSerMetGlnThrThrGluAlaI 313

187 TCTTTTTGTCTTCAAAC.................AGCACTTGGCCTTTT 156
    ::      |||:::|||                 |||       |||
313 leSerGlySerAlaAsnThrProIleHisAsnGlySerLeuValPro... 328

155 TTGACGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATA 106
       |||:::|||   :::|||:::                  ||||||
329 ...ThrThrSerSerAlaLeuVal..................ProThrTh 338

105 TTCTTCGCCAAGCAACGCGACTTCGGTAATGGCCGGGCCGTCA 63
     |||:::        :::::::::|||:::|||::::::  |||
338 rSerAlaAlaHisSerGlyAlaSerAlaMetThrAsnSerSer 352
```

seq_name: SwissProt_40:MURF_HAEIN

```
seq_documentation_block:
ID   MURF_HAEIN     STANDARD;      PRT;   457 AA.
AC   P45061;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-
DE   alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide synthetase)
DE   (D-alanyl-D-alanine-adding enzyme).
GN   MURF OR HI1134.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC       STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC       PRECURSOR OF MUREIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC       glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine = ADP
CC       + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC       carboxy-L-lysyl-D-alanyl-D-alanine.
CC   -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32793; AAC22789.1; -.
DR   HSSP; P11880; 1GG4.
DR   TIGR; HI1134; -.
DR   InterPro; IPR000713; Mur_ligase.
DR   InterPro; IPR004101; Mur_ligase_C.
DR   Pfam; PF01225; Mur_ligase; 1.
DR   Pfam; PF02875; Mur_ligase_C; 1.
KW   Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
KW   ATP-binding; Complete proteome.
FT   NP_BIND     109    115       ATP (POTENTIAL).
SQ   SEQUENCE   457 AA;  50073 MW;  095EFE74F27CD28B CRC64;

alignment_scores:
          Quality:  104.00          Length:    397
            Ratio:  0.547             Gaps:     22
 Percent Similarity:  47.859   Percent Identity: 22.166

alignment_block:
 US-09-303-518D-131 x MURF_HAEIN   ..

 Align seg 1/1  to: MURF_HAEIN  from: 1  to: 457

    253 CGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGA 302
         ::::::||||||:::   :::|||          |||:::::::::::::::
     11 GlnIleLeuGlnAlaLysLeuIle.........GlyAspGluAsnValGl 24

    303 GTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAG 352
         :   |||                           |||::::::::::::
     24 nValGlu.......................LysIleAsnThrAsp.... 31

    353 TGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGT 402
              |||:::   ::::::::::::|||:::   |||||||:::    :::
     32 ..ThrArgLysSerValSerAsnSerLeuPhePheAlaLeuLysGlyGlu 47

    403 CCGTTCAGCAAAATCCCTGCCGTAGAT......GCCGAGCCGTTCGCCAT 446
           |||:::            :::|||        ::::::      |||::
     48 LysPheAspAlaHisGlnTyrLeuAspGlnAlaValSerGlnGlyAlaLe 64

    447 CTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCA 496
         :   :::   :::::::::   |||            :::   |||    ::::
     64 uAlaLeuValValGlnGlnGluAsnSerSerIleSerValProGlnLeuV 81

    497 TCATCAAAGAA.........GCCGCCGAAGACTTCAAACGCGGCCTGTTG 537
```

```
         :::::|||:::          |||      ::::::   :::   |||
  81 alValLysAspThrArgIleAlaLeuGlyGluLeuAlaLysTrpLeu... 96

 538 GTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGG 587
        ::::::::::::   |||   :::   :::     ::::::||
  97 ...ArgGluLysIleAsnProArgThrValAlaMetThrGlySerSerGl 112

 588 CGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACATGAATTTG 637
     |       :::   :::|||   :::|||:::|||       |||
 112 yLysThrThrValLysGluMetThrAlaSerIleLeuGlnHis....... 126

 638 GCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAG 687

 126 .................................................. 126

 688 CCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGAT 737
                                     :::      |||
 127 ...............................ThrAlaAlaAspSerGl 132

 738 TGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGG 787
      |||:::         ||||||:::::|||::::::|||::::::      |
 132 uAlaVal......LeuPheThrAsnGlyAsnPheAsnAsnAspIleGlyV 147

 788 TTGCCTTGGGCGGCCTGCAAGTC...AACAAACCGCGCCTCTTGCGTACC 834
     ||    |||       |||::::::    :::|||   |||:::
 147 alProLeuThrLeuLeuArgLeuThrGluLysHisArgPheAlaValIle 163

 835 GTTTTGGGTGCG.............................AAGGTGTC 854
        |||||||||                               |||:::
 164 GluLeuGlyAlaAsnHisGlnAsnGluIleAsnTyrThrThrLysLeuVa 180

 855 TCAACTTACCGCC..................................... 867
      |||    :::|||
 180 lGlnProAsnAlaAlaLeuIleAsnAsnIleAlaProAlaHisLeuGluG 197

 868 .....GGCGAATTG.........GTTGACGCGGACAACCGCGTGATTTCC 903
         |||:::|||         |||:::|||    :::::::::
 197 lyPheGlySerLeuAlaGlyValValGlnAlaLysGlyGluIleTyrArg 213

 904 GGTTCGGTATTGAACGGT...GCGATTGCACAAGGCGCGCATGATTATTT 950
     |||     :::    ||||||    ||||||    ::::::    |||::::::||
 214 GlyLeuThrLysAsnGlyValAlaIleIleAsnAlaGluHisAsnHisLe 230

 951 GGGACGCTACCACAATCAGATTTCC......GTTATCGAAGAAGGCCGCA 994
     |        ::::::::::::||||||      :::|||:::       ::::
 230 uAspIleTrpGlnLysGluIleSerAsnHisAlaIleGlnTyrPheAsnG 247

 995 GCAAAGAG...................................CTG 1005
     ::|||:::
 247 lyLysAspTyrSerAlaLysAsnIleHisHisThrSerGlnGlySerThr 263

1006 TTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCAC 1055
     |||        ::::::||||||         :::     :::||||||
 264 PheThrLeuIleSerProGln...GlyGluIleGluIleThrLeuProTy 279

1056 TCTCGGC...CATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCG 1102
      ||||||    |||    :::||||||    |||          :::    |||:
 279 rLeuGlyGluHisAsnValLysAsnAlaLeuAlaAlaThrAlaLeuAlaM 296

1103 TCAAC...GGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAG... 1146
     ::|||    |||:::      :::    |||       |||      :::
 296 etAsnValGlyAlaThrLeuThrAspValLysAlaGlyLeuGluGlnArg 312

1147 ..............CGCGTAATGCCGTTGGACATCCTGCCTACCTTGCT 1181
                   |||::::::|||:::::::::    |||:::|||||
 313 SerGlnValLysGlyArgLeuPheProIleGlnValThrProAsnLeuLe 329

1182 TTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTT 1231
     ||||   |||                 ||||||    :::|||
 329 uLeuLeuAsp............AspThrTyrAsnAla............. 337

1232 GCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCT 1270
                    ::::::|||  |||||||||||||||
```

```
 338 ..............AsnLysAsp.SerLeuCysAlaAla 345
```

seq_name: SwissProt_40:SC18_SCHPO

```
seq_documentation_block:
ID   SC18_SCHPO     STANDARD;      PRT;   792 AA.
AC   Q9P7Q4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Vesicular-fusion protein SEC18 homolog.
GN   SPAC1834.11C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR VESICLE-MEDIATED TRANSPORT. CATALYZES THE
CC       FUSION OF TRANSPORT VESICLES WITHIN THE GOLGI CISTERNAE. IS ALSO
CC       REQUIRED FOR TRANSPORT FROM THE ENDOPLASMIC RETICLUM TO THE GOLGI
CC       STACK. SEEM TO FUNCTION AS A FUSION PROTEIN REQUIRED FOR THE
CC       DELIVERY OF CARGO PROTEINS TO ALL COMPARTMENTS OF THE GOLGI STACK
CC       INDEPENDENT OF VESICLE ORIGIN (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL157734; CAB75779.1; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003960; AAA_sub.
DR   InterPro; IPR003959; AAA_subfam.
DR   Pfam; PF00004; AAA; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00674; AAA; 1.
KW   Transport; Protein transport; Endoplasmic reticulum; Golgi stack;
KW   ATP-binding; Repeat.
FT   NP_BIND     311    318       ATP (POTENTIAL).
FT   NP_BIND     592    599       ATP (POTENTIAL).
SQ   SEQUENCE   792 AA;  87553 MW;  C983E7ABAC0E80B3 CRC64;
```

```
alignment_scores:
           Quality:  103.50            Length:    459
             Ratio:  0.545               Gaps:     22
 Percent Similarity:  41.394    Percent Identity:  21.569

alignment_block:
 US-09-303-518D-131 x SC18_SCHPO   ..

 Align seg 1/1  to: SC18_SCHPO  from: 1  to: 792

       7 AAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGT 56
          :::|||   |||   ||||||              |||   :::|||:::
     323 GlnIleGlyLysMetLeuAsn.........AlaArgGluProLys..... 334

      57 CATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAAT 106
           |||    :::||||||    |||                    ::::::|
     335 .IleValAsnGlyProGluIleLeu.................AsnLysT 345

     107 ATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAA 156
          ||||||||                         ::::::   |||   |||:::
```

```
345 yrValGly....................GlnSerGluGluAsnValArg 354

157 AAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTAC 206
    |||          ||||||  |||   ::::::
355 Lys.........LeuPheAlaAspAlaGluArg................. 362

207 TGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCG 256
                                  ||||||||:::
363 ......................GluTyrArgAspArgGlyGluGluSerG 372

257 TACTTCAGTCAGTCGTGATT.............................. 276
      |||:::   :::::::::
372 lyLeuHisIleIleIlePheAspGluLeuAspAlaIleCysLysLysArg 388

277 GCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGC 326
    :::   :::|||:::|||               ::::::  |||   :::
389 GlySerSerGlyGlyAspThrGlyValGlyAspGlnValValAsnGlnLe 405

327 GCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAG 376
     |||||||||:::::::::            :::|||::::::    |
405 uLeuAlaLysMetAspGlyValAspGlnLeuAsnAsnIleLeuValIleG 422

377 GCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTA 426
    ||:::
422 lyMet............................................. 423

427 GATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCT 476
                             :::|||     |||
424 ........................ThrAsnArgLysAsp........... 428

477 GGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAAC 526
                     :::|||   ||||||          ::::::
429 ...................MetIleAspGluAlaLeu......LeuArgP 437

527 GCGGCCTGTTGGTATTG.............................AGC 546
       |||   |||   :::                             :::
437 roGlyArgLeuGluValHisMetGluIleSerLeuProAspGluHisGly 453

547 CGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGC........ 588
    ||||||         |||||||||:::   :::   |||:::
454 ArgLeuGlnIleLeuLysIleHisThrSerArgMetAlaSerAsnGlyIl 470

589 .......GCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACATG 631
               ||||||      |||:::   |||::::::   |||   :
470 eLeuGluAsnAspValAspMetGluGluLeuAlaSerLeu...ThrLysA 486

632 AATTTGGCGGCCCGCATCCTGCCGGCTTG......AGTGGCACGCACATT 675
     ::|||:::|||    :::   |||||||||      |||::::::   :::
486 snPheSerGlyAlaGluIleAlaGlyLeuIleLysSerAlaSerSerPhe 502

676 CATTTC............................................ 681
       |||
503 AlaPheTyrArgHisIleLysValGlyThrThrAlaAlaValSerGlyAs 519

682 .ATCGAGCCAGTCGGCGCGAATAAAACCGTGTGG................ 714
     :::|||   :::   :::|||::::::   :::
519 nLeuGluAsnIleLysValAsnArgAsnAspPheLeuAsnAlaLeuSerG 536

714 .................................................. 714

536 luValArgProAlaTyrGlyValSerGluGluGluLeuGluSerArgVal 552

715 .........ACCATCAATTAT...........CAAGACGTGATTGCTAT 743
               ||||||:::             :::::::::|||:::
553 GlnGlyGlyIleIleAsnPheGlyLysHisIleGluGluIleIleThrGl 569

744 CGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGC.......... 783
     |||:::|||||||||             |||:::||||||
569 uGlyLysLeuPheValGlnGlnValLysAsnSerGluArgThrArgLeuV 586

784 .....GTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTG 828
         |||:::      |||   :::      :::|||      |||
586 alSerValLeuLeuSerGlyProIleAlaSerGlyLysThrAlaLeuAla 602

829 CGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGT 878
     |||:::      |||      |||::::::         :::||||||
603 AlaThrIle......AlaLeuGlySerGluPheProPheValLysLeuVa 617

879 TGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTG 928
     |::::|||::::::   |||         |||          |||:::   :::|
617 lSerAlaGluSerMetVal......GlyMetAsnGluAsnAlaArgValA 632

929 CACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTT 978
     ||:::      :::   :::            |||   :::   :::||||||
632 laHisValAsnArgValPheGluAspSerTyrLysSerProLeuSerVal 648

979 ATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCC 1028
     |||         :::   ::::::::::::     ||||||   |||
649 IleValValAspGluIleGluArgIleIleAspTrpVal...ProIleGl 664

1029 GGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACA 1078
       ::::::|||   |||                    :::|||::::::
664 yProArgPheSerAsnThr..................LeuLeuGlnThrL 675

1079 AACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTA 1128
      :::         |||                  |||   |||   ::::::
675 euMetValLeuPheLysLysGlnProProLysGlyHisArgLeuLeuIle 691

1129 CCGATCGGCACTTATGAGCGCGTAATG 1155
              |||   ||||||:::|||
692 LeuAlaThrThrSerGluArgThrMet 700
```

seq_name: SwissProt_40:VGP3_EBV

```
seq_documentation_block:
ID   VGP3_EBV       STANDARD;      PRT;   907 AA.
AC   P03200; P03201;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Envelope glycoprotein GP340 (Membrane antigen) (MA) [Contains:
DE   Glycoprotein GP220].
GN   BLLF1.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
CC   -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC       B-CELLS.
CC   -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC       ENVELOPE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; V01555; CAA24854.1; -.
DR   PIR; A03762; QQBE21.
DR   PIR; A03763; QQBE22.
DR   PIR; S33008; S33008.
KW   Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    195    195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    277    277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    378    378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    386    386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    443    443       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    497    497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    519    519       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    589    589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    610    610       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    656    656       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    683    683       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    701    701       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    735    735       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    746    746       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    755    755       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    780    780       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    815    815       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    858    858       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    888    888       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    502    698       MISSING (IN GP220).
SQ   SEQUENCE   907 AA;  94431 MW;  0750141CBCAC52C9 CRC64;

alignment_scores:
          Quality:  103.50              Length:    507
            Ratio:  0.470                 Gaps:     26
 Percent Similarity:  43.393   Percent Identity:  22.485

alignment_block:
 US-09-303-518D-131/rev x VGP3_EBV   ..

 Align seg 1/1  to: VGP3_EBV  from: 1  to: 907

   1340 CCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCG..... 1296
        ||||||   :::::::::::||||||          :::   |||
    240 ProSerGlyGlyIleLeuThrSerThrSerProValAlaThrProIlePr 256

   1295 ..........TATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGA 1256
                  |||:::|||               |||          :::|
    256 oGlyThrGlyTyrAlaTyrSerLeuArgLeuThrProArgProValSerA 273

   1255 GGTCTTCTTCGTCCAATTCC.......................AAGCAA 1230
        ||      ::::::||||||                        :::
    273 rgPheLeuGlyAsnAsnSerIleLeuTyrValPheTyrSerGlyAsnGly 289

   1229 CCCAAAGCCTGCGCG................................... 1215
        |||||||||   :::
    290 ProLysAlaSerGlyGlyAspTyrCysIleGlnSerAsnIleValPheSe 306

   1214 .......CTGTCGGTATCGCCGACGATTAAATCGCGCAAAAGCAAGGTAG 1172
                :::   :::|||     :::   ::::::   :::   :::
    306 rAspGluIleProAlaSerGlnAspMetProThrAsnThrThrAspIleT 323

   1171 GCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCGATCGGTACCATG 1122
         :::::::::::     |||    |||    |||||||:::    |||
    323 hrTyrValGlyAspAsnAlaThrTyrSer...ValProMetValThrSer 338

   1121 ......GCGCGGTCGCCGCCGTTGACGGCTGTCGTG.............. 1092
              |||::::||||||    :::|||:::::::::
    339 GluAspAlaAsnSerProAsnValThrValThrAlaPheTrpAlaTrpPr 355

   1091 ....AACTTGAAGAGTTTGTTTTTAGGAAATGGCCGAGAGTGGTGCGCG 1046
            |||    ::::::    |||      ||||||      :::      :
    355 oAsnAsnThrGluThrAspPheLysCysLysTrpThrLeuThrSerGlyT 372

   1045 TGATGGAGTATTTGTCCGGCTGC...........GGCGCAACCCAGCCG 1008
        ::               |||||||||          ||||||
    372 hrPro.........SerGlyCysGluAsnIleSerGlyAlaPheAlaSer 385

   1007 AACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTA 958
        |||             |||     :::||||||
    386 Asn............ArgThrPheAspIleThr................. 392

    957 GCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCG 908
                                        ::::::   ::::::|||
    393 ..............................ValSerGlyLeuGlyThrA 399

    907 AACCGGAA.........ATCACGCGGTTGTCCGCGTCAACCAATTCGCCG 867
          |||:::           |||||||||    :::
    399 laProLysThrLeuIleIleThrArgThrAla................. 409

    866 GCGGTAAGTTGAGACACCTTCGCACCCAAAACGGTACGCAAGAGGCGCGG 817
                          |||    |||      |||:::::|||
    410 ...............ThrAsnAlaThrThrThrThrHisLys........ 418

    816 TTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGAC 767
         ::::::      :::    |||:::::||||||   |||
    419 .ValIlePheSerLysAlaProGluSerThrThrThrSer.......... 431

    766 GGCCTGTTACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGATG 717
          |||:::    |||          :::|||        :::
    432 ..ProThrLeuAsnThrThrGlyPheAlaAspProAsn............ 443

    716 GTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCGT 667
                          :::   ||||||      :::          ||
    444 .................ThrThrThrGlyLeuProSerSerThrHisVa 454

    666 GCCACTCAAG.........CCGGCAGGATGCGGGCCGCCAAATTCATGTG 626
        ||||   :::          ||||||:::    ||||||      |||    :
    454 lProThrAsnLeuThrAlaProAlaSerThrGlyProThrValSerThrA 471

    625 TTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTA 576
        ::::::::    :::                   ||||||::::::    |||:::
    471 laAspValThrSerProThrProAlaGlyThrThrSerGlyAlaSerPro 487

    575 CACACA...............TGG......................... 567
          |||                 |||
    488 ValThrProSerProSerProTrpAspAsnGlyThrGluSerLysAlaPr 504

    566 ....ATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGA 521
             :::        |||:::      ::::::|||      |||:::
    504 oAspMetThrSerSerThrSerProValThrThrProThrProAsnAlaT 521

    520 AGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAGCGGA 471
          |||    :::    ::::::          :::      :::|||:::|||
    521 hrSerProThrProAlaValThrThrProThrProAsnAlaThrSerPro 537

    470 TTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTACGGC 421
                :::::::::     |||          |||:::::::::     |||
    538 ThrProAlaValThr...ThrProThrProAsnAlaThrSerProThrLe 553

    420 AGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCTGATT 371
         |||             :::            |||||||||       |||
    553 uGly.........LysThrSerProThrSerAlaValThrThrProThrP 567

    370 GAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTGCCAGCGCTTCA 321
                                  |||::::::       :::      :::||||||
    567 roAsnAlaThrSerProThrLeuGlyLysThrSerProThrSerAlaVal 583

    320 GGTACGTAGCGTTCGAACTCGATTTCG...............TCGTTGCC 286
```

```
        |||           |||:::  |||                 :::  .||
 584 ThrThrProThrProAsnAlaThrSerProThrLeuGlyLysThrSerPr 600

 285 TTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAA 236
     |::::::|||:::||||||         :::      :::|||     :
 600 oThrSerAlaValThrThrProThrProAsnAlaThrGlyProThr...V 616

 235 TAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTC 186
     :::::            |||:::|||:::|||:::|||   |||
 616 alGlyGluThrSerProGlnAlaAsnAlaThrAsnHisThr......... 629

 185 TTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTT 136
        |||:::::::|||           |||    :::|||
 630 ...LeuGlyGlyThrSerProThrProValValThr.............. 640

 135 GATTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGCGA 86
                                  |||:::|||:::|||||||
 641 ..............................SerGlnProLysAsnAlaT 647

  85 CTTCG...GTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTGCCC 39
     |||||   |||   :::|||          :::|||   |||:::
 647 hrSerAlaValThrThrGlyGlnHisAsnIleThrSerSerSerThrSer 663

  38 GCGATGGGCAGATTTAGACCT 18
     :::|||:::   :::||||||
 664 SerMetSer...LeuArgPro 669
```

seq_name: SwissProt_40:TEGU_HCMVA

```
seq_documentation_block:
ID   TEGU_HCMVA     STANDARD;      PRT;  2241 AA.
AC   P16785;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Probable large tegument protein.
GN   UL48.
OS   Human cytomegalovirus (strain AD169).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
OX   NCBI_TaxID=10360;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90269039; PubMed=2161319;
RA   Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA   Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,
RA   Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT   "Analysis of the protein-coding content of the sequence of human
RT   cytomegalovirus strain AD169.";
RL   Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC   -!- FUNCTION: TEGUMENT PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC       EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X17403; CAA35407.1; -.
DR   PIR; S09811; S09811.
KW   Capsid assembly.
SQ   SEQUENCE   2241 AA;  253222 MW;  F88929E9F9D0458E CRC64;

alignment_scores:
           Quality:  103.50          Length:     538
             Ratio:  0.431             Gaps:      29
 Percent Similarity:  44.610   Percent Identity:  22.491
```

```
alignment_block:
 US-09-303-518D-131 x TEGU_HCMVA   ..

 Align seg 1/1  to: TEGU_HCMVA  from: 1  to: 2241

  22 CTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGC.. 69
     |||:::   |||:::   |||      |||            ||||||
 101 LeuSerArgProPheAsnGlyThrThrGluThrCysAspLeuAspGlyTy 117

  70 .......CCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCG 112
            |||:::|||   :::      :::|||        ::::::
 117 rMetCysProGlyIlePheAsp......PheLeuArgTyrAlaHisAlaL 132

 113 GCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGC 162
        ||||||:::      :::        :::::::::  ::::::
 132 ysProArgProThrTyrValLeuValThrValAsnSerLeuAlaArgAla 148

 163 CAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCC 212
       |||:::   ||||||                :::||||||
 149 ValValPheThrGluAspHisMet.........LeuValPheAspProHi 162

 213 GGCTTCAGGCAAAATC......GCCGCTATTCACCGTGGCGAAAAGCGCG 256
      :::|||:::::        ||||||:::::::::   |||
 162 sSerSerAlaGluCysHisAsnAlaAlaValTyrHisCysGluGly.... 177

 257 TACTTCAGTCAGTCGTGATTGCCGTTGAAGGC..............AAC 291
       |||:::::|||:::::::::::   |||                :::
 178 ..LeuHisGlnValLeuMetValLeuThrGlyPheGlyValGlnLeuSer 193

 292 GACGAAATCGAGTTCGAACGCTAC......................... 315
         :::   :::|||
 194 ProAlaPheTyrTyrGluAlaLeuPheLeuTyrMetLeuAspValAlaTh 210

 316 .GTACCTGAAGCG......CTGGCAAAATTGAGCAGCGAAAAAGTGCGCC 358
      ||||||||||||        |||:::|||   |||           |
 210 rValProGluAlaGluIleAlaAlaArgLeuValSerThrTyrArgAspA 227

 359 GCAACCTGATTCAATCAGGCTTATGG..............ACTGCGCTT 393
     ||::::::       :::|||:::                 ||||||
 227 rgAspIleAspLeuThrGlyValValArgGluSerAlaAspThrAlaAla 243

 394 CGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGC....... 436
       ||| :::   :::|||    |||||||||   :::|||:::
 244 ThrThr.ThrThrAlaAlaProSerLeuProProLeuProAspProIleV 260

 437 ..........................................CGTTCG 442
                                                   :::
 260 alAspProGlyCysProProGlyValAlaProSerIleProValTyrAsp 276

 443 CCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACG 492
     |||||||||||   :::    ||||||   :::              |||
 277 ProSerSerSerProLysLysThrProGluLysArgArgLysAspLeuSe 293

 493 GTCATCA...........TCAAAGAAGCCGCCGAAGACTTCAAACGCGG 530
      :::|||              :::||||||||||||:::|||::::::   :
 293 rGlySerLysHisGlyGlyLysLysLysProProSerThrThrSerLysT 310

 531 CCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAG 580
     ::  |||:::   :::         :::         |||   :::::::::
 310 hr..LeuAlaThrAlaSerSerSerProSerAlaIleAlaAlaAlaSerS 326

 581 CAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACAT 630
     :::::::|||      ||||||   ::::::::::::
 326 erSerSerAlaValProProSerTyrSerCysGlyGlu............ 338

 631 GAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCATTT 680
               |||     ||||||         |||   :::    :::::
 339 .........GlyAlaLeuProAla......LeuGlyArgTyrGlnGlnLe 350

 681 CATCGAG.................................CCAGTCG 694
     :::::::                                 |||:::
 350 uValAspGluValGluGlnGluLeuLysAlaLeuThrLeuProProLeuP 367
```

```
 695 GCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATC 744
      ||||||   ::::::|||||||::::::                 |||
 367 roAlaAsnThrSerAlaTrpThrLeuHis...............AlaAla 378

 745 GGA............................................... 747
     |||
 379 GlyThrGluSerGlyAlaAsnAlaAlaThrAlaThrAlaProSerPheAs 395

 748 .CGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGG 796
      :::    |||:::|||   ||||||:::            :::   |||
 395 pGluAlaPheLeuThrAspArgLeuGlnGlnLeuIleIleHisAla.... 410

 797 GCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCG 846
                |||||||:::        ||||||         |||
 411 ...........ValAsnGlnArgSerCysLeuArgArgProCysGlyPro 423

 847 AAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGT 896
     :::   :::         |||   :::          :::   ::::::
 424 GlnSerAlaAlaGlnGlnAlaValArgAlaTyrLeuGlyLeuSerLysLy 440

 897 GATTTCCGGTTCGGTATTGAAC............GGTGCGATTGCACAAG 934
      :::::::::   :::||||||             |||       |||
 440 sLeuAspAlaPheLeuLeuAsnTrpLeuHisHisGlyLeuAspLeuGlnA 457

 935 GCGCGCATGATTATTTG................................. 951
           |||||||||||
 457 rgMetHisAspTyrLeuSerHisLysThrThrLysGlyThrTyrSerThr 473

 952 ..............................................GG 953
                                                   ||
 474 LeuAspArgAlaLeuLeuGluLysMetGlnValValPheAspProTyrGl 490

 954 ACGCTACCAC...........................AATCAGATTTCCG 976
     ||||   |||                           :::::::::
 490 yArgGlnHisGlyProAlaLeuIleAlaTrpValGluGluMetLeuArgT 507

 977 TTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAG 1026
      :::|||:::        ::::::|||||| |||        |||:::|||
 507 yrValGluSerLysProThrAsnGluLeu.SerGlnArgLeuGlnArgPh 523

1027 CCGGACAAATACTCCA...............TCACGCGCACCACTCTCGG 1061
        |||:::   |||                |||   :::   ||| |||
 523 eValThrLysArgProMetProValSerAspSerPheValCysLeu.Arg 539

1062 CCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGG.. 1109
     |||          ::::::      ||||||   ::::::|||:::|||
 540 ProValAspPheGlnArgLeuThrGlnValIleGluGlnArgArgArgVa 556

1110 ............CGGCGACCGCGCCATGGTACCGATCGGCACTTATGAG 1146
                 |||:::      |||||||:::   :::||||||
 556 lLeuGlnArgGlnArgGluGluTyrHisGlyValTyrGluHisLeu.... 571

1147 CGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAAT 1196
                :::||| ::::::   ::::::   ::::::||||||
 572 ...........AlaGly.LeuIleThrSerIleAspIleHisAspLeuAs 584

1197 CGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAA...TTGG 1243
      ::::::|||    :::       :::    |||   :::|||:::   ||||
 584 pAlaSerAspLeuAsnArgArgGluIleLeuLysAlaLeuGlnProLeuA 601

1244 ACGAA 1248
      ||:::
 601 spAsp 602
```

seq_name: SwissProt_40:SGS3_DROER

```
seq_documentation_block:
ID   SGS3_DROER     STANDARD;      PRT;   328 AA.
AC   P13730;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Salivary glue protein Sgs-3 precursor.
GN   SGS3.
OS   Drosophila erecta (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7220;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88332966; PubMed=3138416;
RA   Martin C.H., Mayeda C.A., Meyerowitz E.M.;
RT   "Evolution and expression of the Sgs-3 glue gene of Drosophila.";
RL   J. Mol. Biol. 201:273-287(1988).
CC   -!- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M14652; -; NOT_ANNOTATED_CDS.
DR   PIR; S01359; S01359.
DR   FlyBase; FBgn0012268; Dere\Sgs3.
KW   Repeat; Signal.
FT   SIGNAL        1     23
FT   CHAIN        24    328       SALIVARY GLUE PROTEIN SGS-3.
SQ   SEQUENCE   328 AA;  36355 MW;  62F27F188C0F8272 CRC64;

alignment_scores:
             Quality:  102.00              Length:    399
               Ratio:  0.729                 Gaps:     14
  Percent Similarity:  35.088   Percent Identity:  20.050

alignment_block:
  US-09-303-518D-131 x SGS3_DROER   ..

  Align seg 1/1  to: SGS3_DROER  from: 1  to: 328

      29 TGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGG..... 73
           ||||||:::|||:::            |||         ||||||||
      28 CysProLysArgThr...ThrProLysProCysThrThrAlaArgProTh 43

      74 .......CCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCAT 116
                |||:::      :::   |||         :::   :::    |
      43 rCysAlaProValThrThrThrThrCysArgProProThrThrThrArgC 60

     117 GCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAG 166
          ||    |||        :::|||       |||:::::::   |||:::
      60 ysProProProThrThrThrArgCysProProProThrArgProAlaGlu 76

     167 TGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCT 216
          |||       |||||||||      |||         ||||||
      77 CysThrAlaThrThrLysArgProThrAlaArgProThrThrArgArgTh 93

     217 TCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTC 266
            :::::::::
      93 rThrValArgAla..................................... 97

     267 AGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACG 316
                        :::||||||||||:::   :::::::::   :::
      98 ........ThrThrLysArgAlaThrThrArgArgThrThrLysArgAla 111

     317 TACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTG 366
            :::|||          |||:::   |||   |||:::|||
     112 ThrThrArgArgThrThrValArgAlaThrThrLysArgAlaThrThr.. 127

     367 ATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAAT 416
               :::          :::   ||||||:::|||    :::    :
     128 .......ArgArgThrThrThrLysArgAlaProThrArgArgThrThrT 142
```

```
 417 CCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACA 466
      ::                         |||:::|||:::          |||   |||
 142 hrLysArgAlaThrThrArgArgAsnProThr.........ArgArgThr 155

 467 CCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAA 516
            |||       ||||||    |||::::::::|||:::
 156 ThrThrArgArgAlaProThrLysArgAlaThrThrLysArgAlaThrTh 172

 517 GACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAAT 566
          |||    :::
 172 rArgArgAsnProThr.................................. 177

 567 CCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCA 616
                          :::   :::|||   |||
 178 .................LysArgLysThrThrArg............... 183

 617 ATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGC 666

 183 .................................................. 183

 667 ACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGAC 716
      ||||||    :::
 184 .ArgThrThrVal..................................... 187

 717 CATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCC 766

 187 .................................................. 187

 767 GTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAA 816
                                      |||    |||:::|||::
 188 .............................ArgAlaThrLysThrThrLy 194

 817 CCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGC 866
     :||||||
 194 sArgAla........................................... 196

 867 CGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGA 916
                          ||||||        |||::::::
 197 ...................ThrThrLysArgAlaProThrLys...... 204

 917 ACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAAT 966
                                           |||||||||
 205 ..................................ArgAlaThrThrLy 209

 967 CAGATTTCCG...........TTATCGAAGAAGGCCGCAGCAAAGAGCT 1004
      |||    |||            :::|||:::|||    :::|||    :
 209 sArgAlaProThrLysArgValThrThrLysArgAlaProThrLysArgA 226

1005 GTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCA 1054
     ::::::::               :::::      |||    :::|||||||||
 226 laThrThrLysArgAlaProThrLysArgAlaThrThrLysArgAlaPro 242

1055 CTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTC 1104
        :::                    :::::::::::::|||    |||::
 243 ThrLys..................ArgAlaThrThrLysArgAlaProTh 253

1105 AACGGCGGCGACC.........GCGCCATGGTACCGATCGGCACTTATGA 1145
     :          ||||||          ||||||       |||::::::
 253 rLysArgAlaThrThrLysArgAlaProThrLysArgAlaThrThrLysA 270

1146 GCGCGTAATGCCGTTGGACATCC...........TGCCTACCTTGC 1180
        |||    :::|||    ||||||              |||    ||||||
 270 rgAlaThrAlaArgProThrSerLysProCysGlyCysLysProCys 285
```

seq_name: SwissProt_40:ATPB_NICPL

```
seq_documentation_block:
ID   ATPB_NICPL     STANDARD;      PRT;   498 AA.
AC   P26529;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ATP synthase beta chain (EC 3.6.3.14).
GN   ATPB.
OS   Nicotiana plumbaginifolia (Leadwort-leaved tobacco), and
OS   Nicotiana bigelovii (Bigelov's tobacco).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4092, 4088;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92390714; PubMed=1387730;
RA   Avni A., Anderson J.D., Holland N., Rochaix J.-D., Gromet-Elhanan Z.,
RA   Edelman M.;
RT   "Tentoxin sensitivity of chloroplasts determined by codon 83 of beta
RT   subunit of proton-ATPase.";
RL   Science 257:1245-1247(1992).
CC   -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC       GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC       SUBUNIT.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
CC   -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X61320; CAA43613.1; -.
DR   EMBL; X61316; CAA43609.1; -.
DR   PIR; S15723; PWNTBC.
DR   PIR; S15722; PWNTBB.
DR   HSSP; P00829; 1BMF.
DR   Mendel; 2235; NICbi;atpB;1.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR004100; ATP-synt_ab_N.
DR   InterPro; IPR000793; ATPase_AB_C.
DR   InterPro; IPR000194; ATPase_alpha_beta.
DR   Pfam; PF00006; ATP-synt_ab; 1.
DR   Pfam; PF00306; ATP-synt_ab_C; 1.
DR   Pfam; PF02874; ATP-synt_ab_N; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW   ATP synthesis; Chloroplast; Thylakoid; Membrane; CF(1);
KW   Hydrolase; ATP-binding; Hydrogen ion transport.
FT   NP_BIND     172    179       ATP (BY SIMILARITY).
SQ   SEQUENCE   498 AA;  53525 MW;  94ADC63E8BCBF5B8 CRC64;
```

```
alignment_scores:
          Quality:  102.00           Length:    492
            Ratio:  0.493              Gaps:     28
 Percent Similarity:  42.073   Percent Identity: 22.154

alignment_block:
 US-09-303-518D-131 x ATPB_NICPL   ..

 Align seg 1/1  to: ATPB_NICPL  from: 1  to: 498

   178 GACAAAAAGAATCCGGGC........................GTAGTATT 203
        :::||||||||||||||||                        |||:::
    16 GluLysLysAsnProGlyArgValValGlnIleIleGlyProValLeuAs 32

   204 TACTGCGCCGGCTTCAGGCAAAATC...............GCCGCTATTC 238
        :::|||          ||||||:::               |||    :::
```

```
  32 pValAlaPheProProGlyLysMetProAsnIleTyrAsnAlaLeuValV 49

 239 ACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAA... 285
        :::|||:::        |||   |||   :::   :::|||   |||
  49 alGlnGlyArgAspSerValGlyGlnProIleAsnValAlaCysGluVal 65

 286 ...........GGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGA 323
                |||||:::::::::                     ::
  66 GlnGlnLeuLeuGlyAsnAsnArgVal.....................Ar 75

 324 AGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTG......A 367
     :|||:::|||      :::::::::   :::   |||::::::      |
  75 gAlaValAlaMetSerAlaThrAspGlyLeuThrArgGlyMetGluValI 92

 368 TTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATC 417
      ||:::::|||                     |||:::|||   :::
  92 leAspThrGly.....................AlaProIleSer...Val 100

 418 CCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC................. 450
      |||:::     |||          ||||||
 101 ProValGlyGlyAlaThrLeuGlyArgIlePheAsnValLeuGlyGluPr 117

 451 ..................GTCAATGCGATGGACACCAATCCGCTG.... 477
                       |||::::::     |||:::|||:::
 117 oValAspAsnLeuGlyProValAspThrSerThrThrSerProIleHisA 134

 478 ..GCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAA 525
       :::|||      :::   :::              :::   |||:::
 134 rgSerAlaProAlaPheIleGlnLeuAspThrLysLeuSerIlePheGlu 150

 526 CGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTG 575
        |||:::   |||::::::   |||
 151 ThrGlyIleLysValValAspLeuLeu....................... 159

 576 TAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAA 625
                       :::|||   :::::::::::::::|||
 160 ..................AlaProTyrArgArgGlyGlyLysIleGlyL 170

 626 CACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATT 675
              |||||||||
 170 eu......PheGlyGly................................ 173

 676 CATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTA 725
                    :::|||::::::|||||||||     ::::::
 174 ...............AlaGlyValGlyLysThrValLeuIleMetGluLe 185

 726 TCAAGACGTGATTGCT..................................A 742
        :::   ||||||                                   :
 185 uIleAsnAsnIleAlaLysAlaHisGlyGlyValSerValPheGlyGlyV 202

 743 TCGGACGT.....................TTGTTCGTA............ 759
     ::|||:::                     |||::::::
 202 alGlyGluArgThrArgGluGlyAsnAspLeuTyrMetGluMetLysGlu 218

 760 ACAGGCCGTCTGAATACCGAGCGCGTG...........GTTGCCTTGGG 797
     :::|||   :::|||   |||::::::           |||||||||
 219 SerGlyValIleAsnGluGluAsnIleAlaGluSerLysValAlaLeuVa 235

 798 CGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGA 847
          |||:::|||:::|||         |||     :::|||
 235 lTyrGlyGlnMetAsnGluProProGlyAlaArgMetArgValGly.... 250

 848 AGGTGTCTCAACTTACCGCCGGCGAA........................ 873
       ::::::   ||||||   :::|||
 251 ..LeuThrAlaLeuThrMetAlaGluTyrPheArgAspValAsnGluGln 266

 874 ......TTGGTTGACGCGGACAAC.........CGCGTGATTTCCGGTTC 908
           |||:::      ||||||         |||   :::||||||
 267 AspValLeuLeuPheIleAspAsnIlePheArgPheValGlnAlaGlySe 283

 909 GGTATTGAACGGTGCGATTGCACAA......GGCGCGCATGATTAT.... 948
      |  :::::::::   :::::::::      :::|||   |||
 283 rGluValSerAlaLeuLeuGlyArgMetProSerAlaValGlyTyrGlnP 300

 949 ................TTGGGACGCTACCACAATCAGATTTCCGTTATC 981
                     :::|||   :::::::::|||:::
 300 roThrLeuSerThrGluMetGlySerLeuGlnGluArgIleThrSerThr 316

 982 GAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGA 1031
     :::||||||      ::::::   :::      :::|||      ||
 317 LysGluGlySerIleThrSerIleGlnAlaValTyrValProAlaAspAs 333

1032 CAAATACTCC...ATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACA 1078
     |        :::         ||||||::::::|||
 333 pLeuThrAspProAlaProAlaThrThrPheAlaHis............ 345

1079 AACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTA 1128
        :::      ||||||:::::::::         |||:::::::::
 346 .....LeuAspAlaThrThrValLeuSer.........ArgGlyLeuAla 357

1129 CCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTT 1178
        |||   |||      |||   |||||||||         |||
 358 AlaLysGlyIleTyrProAlaValAspProLeuAspSerThrSerThr.. 373

1179 GCTTTTGCGCGATTTAATCGTCGGCGAT...................... 1206
      :::|||:::      |||||||||:::
 374 .MetLeuGlnProArgIleValGlyGluGluHisTyrGluThrAlaGlnA 390

1207 ..................................ACCGACAGCGCGCAG 1221
                                       |||
 390 rgValLysGlnThrLeuGlnArgTyrLysGluLeuGlnAspIleIleAla 406

1222 GCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGC.. 1269
        ||||||      ||||||:::|||||||||      |||   :::
 407 IleLeuGlyLeuAspGluLeuSerGluGluAspArgLeuLeuValAlaAr 423

1270 ..................................TTCGTCTGC.... 1278
                                       ||||||:::
 423 gAlaArgLysIleGluArgPheLeuSerGlnProPhePheValAlaGluV 440

1279 .............CCGGGCAAATAC 1290
                  ||||||||||||
 440 alPheThrGlySerProGlyLysTyr 448
```

seq_name: SwissProt_40:HUTH_BACSU

seq_documentation_block:

```
ID   HUTH_BACSU     STANDARD;      PRT;   508 AA.
AC   P10944;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN   HUTH.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88257040; PubMed=2454913;
RA   Oda M., Sugishita A., Furukawa K.;
RT   "Cloning and nucleotide sequences of histidase and regulatory genes
RT   in the Bacillus subtilis hut operon and positive regulation of the
RT   operon.";
RL   J. Bacteriol. 170:3199-3205(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=95219088; PubMed=7704263;
RA   Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT   "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT   genome containing the hut and wapA loci.";
RL   Microbiology 141:337-343(1995).
CC   -!- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC   -!- PATHWAY: Histidine degradation; first step.
```

```
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- INDUCTION: BY HISTIDINE.
CC   -!- PTM: CONTAINS AN ACTIVE SITE 4-METHYLIDENE-IMIDAZOLE-5-ONE (MIO),
CC       WHICH IS FORMED AUTOCATALYTICALLY BY CYCLIZATION AND DEHYDRATION
CC       OF RESIDUES ALA-SER-GLY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PAL / HISTIDASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M20659; AAA22538.1; -.
DR   EMBL; D31856; BAA06644.1; -.
DR   EMBL; Z99124; CAB15971.1; -.
DR   PIR; S18810; UFBSHS.
DR   HSSP; P21310; 1B8F.
DR   SubtiList; BG10667; hutH.
DR   InterPro; IPR001106; PAL.
DR   Pfam; PF00221; PAL; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Histidine metabolism; Complete proteome.
FT   SITE        141    143       MODIFIED TO FORM 4-METHYLIDENE-IMIDAZOLE-
FT                                5-ONE (BY SIMILARITY).
SQ   SEQUENCE   508 AA;  55674 MW;  869C323BFCC318E0 CRC64;

alignment_scores:
           Quality:  101.50             Length:    320
             Ratio:  0.668                Gaps:     18
 Percent Similarity:  47.500   Percent Identity: 23.750

alignment_block:
 US-09-303-518D-131 x HUTH_BACSU   ..

 Align seg 1/1  to: HUTH_BACSU  from: 1  to: 508

      22 CTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGC.. 69
         |||   |||   :::   |||:::   |||   |||:::::::::|||
     151 LeuAlaLeuAlaLeuIleGlyGlnGlyGlu...ValPhePheGluGlyGl 166

      70 .......CCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCG 112
                |||||||:::|||                  |||   :::   :::|
     166 uArgMetProAlaMetThr...............GlyLeuLysLysAlaG 178

     113 GCATGCGCCCC...TCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAA 159
          ||:::::|||   ::::::      |||||||||   |||:::   :::
     178 lyIleGlnProValThrLeuThrSerLysGluGlyLeuAlaLeuIleAsn 194

     160 GGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGC 209
         |||                                   :::   :::
     195 Gly..................................ThrGlnAlaMe 199

     210 GCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTAC 259
            |||   |||   :::|||   |||   ::::::||||||   :::
     199 tThrAlaMetGlyValValAlaTyrIle...GluAlaGluLysLeuAlaT 215

     260 TTCAGTCAGTCGTGATTGCC.........GTTGAAGGC............ 288
           |||:::      ||||||          :::||||||
     215 yrGlnThrGluArgIleAlaSerLeuThrIleGluGlyLeuGlnGlyIle 231

     289 ...........AACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGC 326
                    :::::::::|||::::::   |||         |||
     232 IleAspAlaPheAspGluAspIleHisLeuAlaArgGlyTyrGlnGluGl 248

     327 GCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAG 376
          :::               :::|||::::::|||      |||   :::|||
     248 nIleAsp......ValAlaGluArgIleArgPheTyrLeuSerAspSerG 263

     377 GCTTATGGACTGCG.........CTTCGCACCCGT.........CCGTTC 408
          |||||   |||:::            ||||||::::::         :::
     263 lyLeuThrThrSerGlnGlyGluLeuArgValGlnAspAlaTyrSerLeu 279

     409 AGCAAAATCCCTGCCGTAGATGCC.......................... 432
              ||||||   |||   :::
     280 ArgCysIleProGlnValHisGlyAlaThrTrpGlnThrLeuGlyTyrVa 296

     433 ....GAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGG 478
             |||   :::   |||   :::||||||        |||||||||
     296 lLysGluLysLeuGluIleGluMetAsnAlaAlaThrAspAsnProLeuI 313

     479 CT.........GCCGACCCTACGGTCATC..................... 498
                    :::|||   ::::::
     313 lePheAsnAspGlyAspLysValIleSerGlyGlyAsnPheHisGlyGln 329

     499 ...ATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAG 545
            |||      |||   :::   :::|||   ::::::      |||::
     330 ProIleAlaPheAlaMetAspPheLeuLysIleAlaIleSerGluLeuAl 346

     546 CCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACG 595
         :::::::::||||||:::|||
     346 aAsnIleAlaGluArgArgIle........................... 353

     596 TGCCGTCTGAAAATGCTGCCAATATCGAAACACATGAA.........TTT 636
                   |||:::   :::|||   :::   ::::::           |||
     354 ........GluArgLeuValAsnProGlnLeuAsnAspLeuProProPhe 367

     637 GGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATT........... 675
            :::||||||||||:::      ||||||:::   |||
     368 LeuSerProHisProGlyLeuGlnSerGlyAlaMetIleMetGlnTyrAl 384

     676 ...........................CATTTCATCGAGCCAGTCGGCG 697
                                   :::   :::|||::::::
     384 aAlaAlaSerLeuValSerGluAsnLysThrLeuAlaHisProAlaSerV 401

     698 CGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGA 747
          ::::::::         :::   |||   :::|||   :::::::::|||
     401 alAspSerIleProSerSerAlaAsnGlnGluAspHisValSerMetGly 417

     748 CGTTTGTTCGTAACAGGCCGT...........CTGAATACCGAGCGCGT 785
            :::   :::                            ||||||:::|||||
     418 ThrIleAlaAlaArgHisAlaTyrGlnValIleAlaAsnThrArgArgVa 434

     786 GGTTGCCTTG 795
          |:::|||:::
     434 lIleAlaIle 437
```